THE ANNALS OF MATHEMATICAL STATISTICS

VOL. II AUGUST, 1931 NO. 3

CONTENTS

PUBLISHED QUARTERLY BY
AMERICAN STATISTICAL ASSOCIATION

Publication Office—Edwards Brothers, Inc., Ann Arbor, Michigan
Business Office—530 Commerce Bldg., New York Univ., New York, N. Y.

Entered as second class matter at the Postoffice at Ann Arbor, Mich., under the Act of March 3rd, 1879.

CORRECTION FOR THE MOMENTS OF A FREQUENCY DISTRIBUTION IN TWO VARIABLES[1]

By

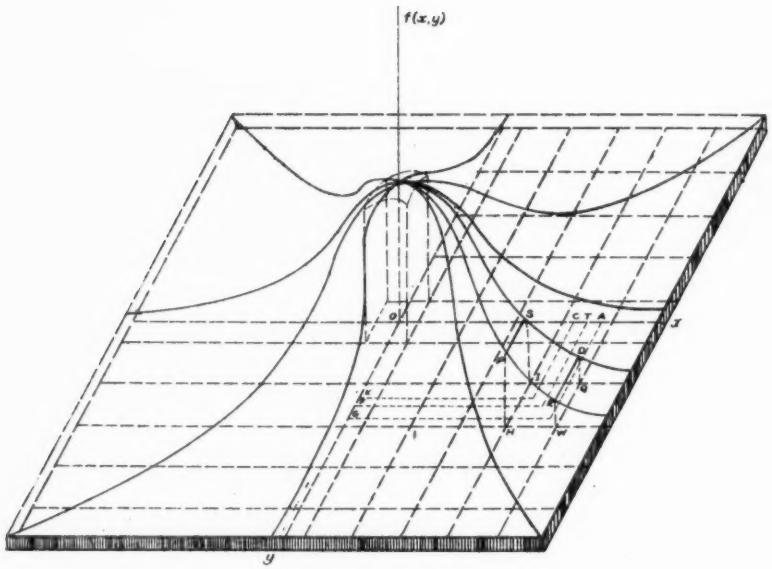
WILLIAM DOWELL BATEN
University of Michigan

In certain statistical problems it is beneficial to divide the given data into classes or groups and investigate the distribution in this form. The moments determined for the distribution divided into classes differ from the moments determined from the original data. It is the object of this article to show how to modify the former to secure the latter for a frequency distribution in two variables.

After the data, given for a frequency distribution of one variable, have been divided into classes the class mark is then the representive of the items in a class. This is assuming that the mean of the items falling in a class is equal to the class mark. For a large number of items in a class, distributed throughout the entire class, the class mark differs very little from the average of the items in the class. But the average of the items raised to a power is not equal to the class mark raised to the same power. Hence corrections should be made to the moments determined from a distribution which is divided into classes.

For a distribution of two variables x and y the data are divided into xy-classes, where the class mark of an x-class

[1]Presented to the American Mathematical Society, Sept. 12, 1930.


f(x,y)
y

is considered to be the representive of the items falling in this class, while the class mark of a y -class is the representive of all items in this particular class. The coordinates of the point in the xy -plane, whose abscissa is the class mark of the x -class and whose ordinate is the class mark of the y -class, may be considered to be the class mark of the double class or the xy -class.

Let the frequencies of the distribution be represented by the volumes of the volume-compartments as shown in the figure. The sum of all such compartments is the total of the frequencies and should be equal to the number of items in the distribution. The little solid $HWQI\text{-}SPRD$ is the frequency of the items falling in the 5th x -class and in the 3rd y -class. OT and OF are the class marks of this x-class and this y -class. $(OT)^n(OF)^m$ multiplied by the frequency of the items falling in this double xy-class may differ considerably from the sum $(OC)^n(OK)^m + (OA)^n(OG)^m + \ldots$, hence corrections must be made to the moments obtained from the distribution divided into classes where the double class marks are the representives of the items in the class. If the class units are made smaller and are allowed to become very near to zero the errors are not so large, for it must be remembered that our results are only approximations.

By definition the n 'th, m'th moment of the distribution which is divided into classes is

$$\nu'_{n:m} = \frac{\Sigma\Sigma x_i^n y_j^m}{N} \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} f(x_i + h, y_j + k)\, dh\, dk,$$

where (x_i, y_j) is considered to be the class mark of the i, j -class, and the double summation extends over all the classes. It is further assumed that $f(x_i + h, y_j + k)$ is such a function which can be expanded into a Taylor series. The above becomes

$$\frac{\Sigma\Sigma x_i^n y_j^m}{N}\int_{-\frac{1}{2}}^{\frac{1}{2}}\int_{-\frac{1}{2}}^{\frac{1}{2}} f(x_i+h, y_j+k)\,dh\,dk = \frac{\Sigma\Sigma x_i^n y_j^m}{N}\int_{-\frac{1}{2}}^{\frac{1}{2}}\int_{-\frac{1}{2}}^{\frac{1}{2}}\Big\{f(x_i, y_j)$$

$$+h\frac{\partial f(x_i,y_j)}{\partial x_i}+k\frac{\partial f(x_i,y_j)}{\partial y_j}+\frac{1}{2!}\left[h^2\frac{\partial^2 f(x_i,y_j)}{\partial x_i^2}+\frac{2hk\,\partial^2 f(x_i,y_j)}{\partial x_i\,\partial y_j}\right.$$

$$\left.+\frac{k^2\partial^2 f(x_i,y_j)}{\partial y_j^2}\right]+\frac{1}{3!}\left[\frac{h^3\partial^3 f(x_i,y_j)}{\partial x_i^3}+\frac{3h^2k\,\partial^3 f(x_i,y_j)}{\partial x_i^2\partial y_j}+\frac{3hk^2\partial^3 f(x_i,y_j)}{\partial x_i\,\partial y_j^2}\right.$$

$$\left.+\frac{k^3\partial^3 f(x_i,y_j)}{\partial y_j^3}\right]+\cdots\Big\}dh\,dk=\frac{\Sigma\Sigma x_i^n y_j^m}{N}\Big\{f(x_i,y_j)$$

$$+\left[\frac{\partial^2 f(x_i,y_j)}{2^2 3!\,\partial x_i^2}+\frac{\partial^2 f(x_i,y_j)}{2^2 3!\,\partial y_j^2}\right]+\left[\frac{\partial^4 f(x_i,y_j)}{2^4 5!\,\partial x_i^4}+\frac{2\,\partial^4 f(x_i,y_j)}{2^4 3\cdot 5!\,\partial x_i^2\partial y_j^2}\right.$$

$$\left.+\frac{\partial^4 f(x_i,y_j)}{2^4 5!\,\partial y_j^4}\right]+\frac{1}{6!}\left[\frac{\partial^6 f(x_i y_j}{2^6\cdot 7\;\partial x_i^6}+\frac{\partial^6 f(x_i,y_j)}{2^6\,\partial x_i^4\,\partial y_j^2}\right.$$

$$\left.+\frac{\partial^6 f(x_i,y_j)}{2^6\,\partial x_i^2\,\partial y_j^4}+\frac{\partial^6 f(x_i,y_j)}{2^6\cdot 7\,\partial y_j^6}\right]+\cdots\Big\}.$$

Now use the Euler-Maclaurin Summation* formula for two variables for finding the value of this double summation. This formula is

*This formula is developed on pages 317-319.

$$\sum_{c}^{d}\sum_{a}^{b} U(x,y)=\int_{c}^{d+1}\int_{0}^{b+1} U(x,y)\,dx\,dy-\frac{1}{2}\int_{c}^{d+1} U(x,y)\,dy\Big]_{a}^{b+1}-\frac{1}{2}\int_{a}^{b+1} U(x,y)\,dx\Big]_{a}^{d+1}$$

$$+\frac{1}{12}\frac{\partial}{\partial y}\int_{a}^{b+1} U(x,y)\,dx\Big]_{c}^{d+1}+\frac{1}{12}\frac{\partial}{\partial x}\int_{c}^{d+1} U(x,y)\,dy\Big]_{a}^{b+1}+\frac{U(x,y)}{4}\Big]_{c}^{d+1}\Big]_{a}^{b+1}-\frac{\partial U(x,y)}{24\,\partial x}\Big]_{c}^{d+1}\Big]_{a}^{b+1}$$

$$-\frac{\partial U(x,y)}{24\,\partial y}\Big]_{c}^{d+1}\Big]_{a}^{b+1}-\frac{\partial^3}{720\,\partial x^3}\int_{c}^{d+1} U(x,y)\,dy\Big]_{a}^{b+1}-\frac{\partial^3}{720\,\partial y^3}\int_{a}^{b+1} U(x,y)\,dx\Big]_{c}^{d+1}$$

$$+\frac{\partial^2 U(x,y)}{144\,\partial x\,\partial y}\Big]_{c}^{d+1}\Big]_{a}^{b+1}+\frac{\partial^3 U(x,y)}{1440\,\partial x^3}\Big]_{c}^{d+1}\Big]_{a}^{b+1}+\frac{\partial^3 U(x,y)}{1440\,\partial y^3}\Big]_{c}^{d+1}\Big]_{a}^{b+1}+\cdots$$

which is the double summation of the function $U(x,y)$ from a to b on the x-axis and from c to d along the y-axis. Applying this formula to the double summation above

$$V'_{n:m}=\iint x^n y^m\Big[f(x,y)+\frac{1}{2!2^2}\Big[\frac{\partial^2 f(x,y)}{3\,\partial x^2}+\frac{\partial^2 f(x,y)}{3\,\partial y^2}\Big]$$

$$+\frac{1}{4!2^4}\Big[\frac{\partial^4 f(x,y)}{5\,\partial x^4}+\frac{\binom{4}{2}\partial^4 f(x,y)}{3\cdot 3\,\partial x^2\,\partial y^2}+\frac{\partial^4 f(x,y)}{5\,\partial y^4}\Big]$$

$$+\frac{1}{6!2^6}\left[\frac{\partial^6 f(x,y)}{7\,\partial x^6}+\frac{\binom{6}{2}\partial^6 f(x,y)}{5\cdot 3\,\partial x^4\partial y^2}+\frac{\binom{6}{4}\partial^6 f(x,y)}{3\cdot 5\,\partial x^2\partial y^4}+\frac{\partial^6 f(x,y)}{7\,\partial y^6}\right]$$

$$+\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$+\frac{1}{s!2^s}\left[\frac{\partial^s f(x,y)}{(s+1)\partial x^s}+\frac{\binom{s}{2}\partial^s f(x,y)}{(s-2+1)(2+1)\partial x^{s-2}\partial y^2}+\frac{\binom{s}{4}\partial^s f(x,y)}{(s-4+1)(4+1)\partial x^{s-4}\partial y^4}\right.$$

$$+\frac{\binom{s}{6}\partial^s f(x,y)}{(s-6+1)(6+1)\partial x^{s-6}\partial y^6}+\cdots+\frac{\binom{s}{t}\partial^s f(x,y)}{(s-t+1)(t+1)\partial x^{s-t}\partial y^t}$$

$$\left.+\cdots\cdots\cdots\cdots+\frac{\partial^s f(x,y)}{(s+1)\,\partial y^s}\right]\Bigg\}dxdy+0+0+\cdots;$$

t is an even number. In obtaining this result it was assumed that $f(x,y)$, $f'(x,y)$, $x^k y^w f(x,y)$, $x^k y^w f'(x,y)$ vanish or become negligible at the limits on the x and y axes, k and w are positive integers.

Therefore

$$\nu'_{n:m}=\mu'_{n:m}+\frac{2!}{2^2 3!}\binom{n}{2}\mu'_{n-2:m}+\frac{2!}{2^2 3!}\binom{m}{2}\mu'_{n:m-2}$$

$$+\frac{1}{4!2^4}\left[\frac{4!}{5}\binom{n}{4}\mu'_{n-4:m}+\frac{(2!)^2}{3\cdot 3}\binom{4}{4}\binom{n}{2}\binom{m}{2}\mu'_{n-2:m-2}+\frac{4!}{5}\binom{m}{4}\mu'_{n:m-4}\right.$$

$$+\frac{1}{6!2^6}\left[\frac{6!}{7}\binom{n}{6}\mu_{n-4:m}+\frac{4!2!}{5\cdot 3}\binom{6}{2}\binom{n}{4}\binom{m}{2}\mu'_{n-4:m-2}+\frac{2!4!}{3\cdot 5}\binom{6}{4}\binom{n}{2}\binom{m}{4}\mu_{n-2:m-4}\right.$$

$$\left.\left.+\frac{6!}{7}\binom{m}{6}\mu'_{n:m-6}\right]+\cdots\cdots\cdots\cdots\cdots\cdots\cdots\right.$$

$$+\frac{1}{s!2^s}\left[\frac{s!}{(s+1)}\binom{n}{s}\mu'_{n-s:m}+\frac{(s-2)!2!}{(s-2+1)(2+1)}\binom{s}{2}\binom{n}{s-2}\binom{m}{2}\mu'_{n-\overline{s-2}:m-2}\right.$$

$$+\frac{(s-4)!\,4!}{(s-4+1)(5)}\binom{s}{4}\binom{n}{s-4}\binom{m}{4}\mu'_{n-\overline{s-4}:m-4}$$

$$+\frac{(s-6)!\,6!}{(s-6+1)(7)}\binom{s}{6}\binom{n}{s-6}\binom{m}{6}\mu'_{n-\overline{s-6}:m-6}+\cdots$$

$$\cdots+\frac{(s-t)!\,t!}{(s-t+1)(t+1)}\binom{s}{t}\binom{n}{s-t}\binom{m}{t}\mu'_{n-\overline{s-t}:m-t}+\cdots$$

$$+\frac{s!}{(s-1)}\binom{m}{s}\mu'_{n:m-s}\Big]+\cdots\cdots$$

If $m=0$ the formula becomes the formula for obtaining the moments about a fixed origin for one variable. This has been done by Sheppard and Carver.

If n and m take on integral values

$$\nu'_{1:1}=\mu'_{1:1},$$

$$\nu'_{2:1}=\mu'_{2:1}+\frac{1}{12}\mu'_{0:1},\ \nu'_{1:2}=\mu'_{1:2}+\frac{1}{12}\mu'_{1:0},$$

$$\nu'_{2:2}=\mu'_{2:2}+\frac{1}{12}\mu'_{0:2}+\frac{1}{12}\mu'_{2:0}+\frac{1}{144},$$

$$\nu'_{3:1}=\mu'_{3:1}+\frac{1}{4}\mu'_{1:1},\ \nu'_{1:3}=\mu'_{1:3}+\frac{1}{4}\mu'_{1:1},$$

$$\nu'_{3:2}=\mu'_{3:2}+\frac{1}{4}\mu'_{1:2}+\frac{1}{12}\mu'_{3:0}+\frac{1}{48}\mu'_{1:0},$$

$$\nu'_{2:3}=\mu'_{2:3}+\frac{1}{12}\mu'_{0:3}+\frac{1}{4}\mu'_{2:1}+\frac{1}{48}\mu'_{0:1},$$

$$\nu'_{3:3}=\mu'_{3:3}+\frac{1}{4}\mu'_{1:3}+\frac{1}{4}\mu'_{3:1}+\frac{1}{16}\mu'_{1:1},$$

$$\nu'_{4:1}=\mu'_{4:1}+\frac{1}{2}\mu'_{2:1}+\frac{1}{80}\mu'_{0:1},\ \nu'_{1:4}=\mu'_{1:4}+\frac{1}{2}\mu'_{1:2}+\frac{1}{80}\mu'_{1:0}$$

$$\nu'_{4:2}=\mu'_{4:2}+\frac{1}{2}\mu'_{2:2}+\frac{1}{12}\mu'_{4:0}+\frac{1}{80}\mu'_{0:2}+\frac{1}{24}\mu'_{2:0}+\frac{1}{960},$$

$$\nu'_{2:4} = \mu'_{2:4} + \tfrac{1}{12}\mu'_{0:4} + \tfrac{1}{2}\mu'_{2:2} + \tfrac{1}{80}\mu'_{2:0} + \tfrac{1}{24}\mu'_{0:2} + \frac{1}{960},$$

$$\nu'_{4:3} = \mu'_{4:3} + \tfrac{1}{2}\mu'_{2:3} + \tfrac{1}{4}\mu'_{4:1} + \tfrac{1}{80}\mu'_{0:3} + \tfrac{1}{8}\mu'_{2:1} + \tfrac{1}{160}\mu'_{0:1},$$

$$\nu'_{3:4} = \mu'_{3:4} + \tfrac{1}{2}\mu'_{3:2} + \tfrac{1}{4}\mu'_{1:4} + \tfrac{1}{80}\mu'_{3:0} + \tfrac{1}{8}\mu'_{1:2} + \tfrac{1}{160}\mu_{1:0},$$

$$\nu'_{4:4} = \mu'_{4:4} + \tfrac{1}{2}\mu'_{2:4} + \tfrac{1}{2}\mu'_{4:2} + \tfrac{1}{80}\mu'_{0:4} + \tfrac{1}{4}\mu'_{2:2} + \tfrac{1}{80}\mu'_{4:0}$$

$$+ \frac{1}{160}\mu'_{2:0} + \frac{1}{160}\mu'_{0:2} + \frac{1}{6400},$$

. . . . . .

From the above the μ''s can be obtained.

$$\mu'_{1:1} = \nu'_{1:1},$$

$$\mu'_{2:1} = \nu'_{2:1} - \tfrac{1}{12}M_y, \quad \mu'_{1:2} = \nu'_{1:2} - \tfrac{1}{12}M_x,$$

$$\mu'_{3:1} = \nu'_{3:1} - \tfrac{1}{4}\nu'_{1:1}, \quad \mu'_{1:3} = \nu'_{1:3} - \tfrac{1}{4}\nu'_{1:1},$$

$$\mu'_{3:2} = \nu'_{3:2} - \tfrac{1}{4}\nu'_{2:1} - \tfrac{1}{12}\nu'_{3:0}, \quad \mu'_{2:3} = \nu'_{2:3} - \tfrac{1}{4}\nu'_{1:2} - \tfrac{1}{12}\nu'_{0:3},$$

$$\mu'_{3:3} = \nu'_{3:3} - \tfrac{1}{4}\nu'_{3:1} - \tfrac{1}{4}\nu'_{1:3} + \tfrac{1}{16}\nu'_{1:1}, \quad \text{etc.}$$

By translating the origin to (M_x, M_y)

$$\mu_{1:1}=\nu_{1:1},$$

$$\mu_{2:1}=\nu_{2:1},\quad \mu_{1:2}=\nu_{1:2},$$

$$\mu_{2:2}=\nu_{2:2}-\frac{1}{12}(\nu_{0:2}+\nu_{2:0})+\frac{1}{144},$$

$$\mu_{3:1}=\nu_{3:1}-\nu_{1:1},\quad \mu_{1:3}=\nu_{1:3}-\nu_{1:1},$$

$$\mu_{3:2}=\nu_{3:2}-\frac{1}{2}\nu_{2:1}-\frac{1}{12}\nu_{3:0},$$

$$\mu_{2:3}=\nu_{2:3}-\frac{1}{2}\nu_{1:2}-\frac{1}{12}\nu_{0:3},$$

$$\mu_{3:3}=\nu_{3:3}-\frac{1}{4}\nu_{1:3}-\frac{1}{4}\nu_{3:1}+\frac{1}{16}\nu_{1:1},$$

etc.

In making corrections for the double moments it must be remembered to correct the single moments of the x's and the y's.*

Euler-Maclaurin Summation for Two Variables

Suppose it is possible to find a function $g(x,y)$ such that $g(x+1,y+1)-g(x+1,y)-g(x,y+1)+g(x,y)=f(x,y)$, or $\Delta_x\Delta_y g(x,y)=f(x,y)$ or $\Delta_x^{-1}\Delta_y^{-1}f(x,y)=g(x,y)$, where Δ represents finite difference and Δ^{-1} represents finite integration. If $g(x,y)$ is such a function, then

$$g(a+1,c+1)-g(a+1,\ c)-g(a\ ,c+1)+g(a\ ,\ c)=f(a\ ,\ c),$$

$$g(a+2,c+1)-g(a+2,\ c)-g(a+1,c+1)+g(a+1,\ c)=f(a+1,\ c),$$

$$g(a+1,c+2)-g(a+1,c+1)-g(a\ ,c+2)+g(a\ ,c+1)=f(a\ ,c+1),$$

$$g(a+2,c+2)-g(a+2,c+1)-g(a+1,c+2)+g(a+1,c+1)=f(a+1,c+1),$$

*See Frequency Curves by H. C. Carver in Handbook of Math. Statistics.

. . . . . .

. . . . . .

$$g(b,d) - g(b,d-1) - g(b-1,d) + g(b-1,d-1) = f(b-1,c-1),$$

$$g(b+1,d+1) - g(b+1,d) - g(b,d+1) + g(b,d) = f(b,d).$$

Add: $g(b+1,d+1) - g(b+1,c) - g(a,d+1) + g(a,c) = \sum_{c}^{d}\sum_{a}^{b} f(x,y).$

Or
$$\sum_{c}^{d}\sum_{a}^{b} f(x,y) = g(x,y)\Big]_{c}^{d+1}\Big]_{a}^{b+1}$$

If it is possible to find the function $g(x,y)$ then the double sum $\sum_{c}^{d}\sum_{a}^{b} f(x,y)$ can be found. Expand $g(x+1,y+1)$ in a Taylor series.

$$g(x+1,y+1) = g(x,y) + \frac{\partial g}{\partial x} + \frac{\partial g}{\partial y} + \frac{1}{2!}\left[\frac{\partial^2 g}{\partial x^2} + \frac{2\,\partial^2 g}{\partial x\,\partial y} + \frac{\partial^2 g}{\partial y^2}\right]$$

$$+\frac{1}{3!}\left[\frac{\partial^3 g}{\partial x^3} + \frac{3\,\partial^3 g}{\partial x^2 \partial y} + \frac{3\,\partial^3 g}{\partial x \partial y^2} + \frac{\partial^3 g}{\partial y^3}\right] + \cdots$$

$$= (e^{\frac{\partial}{\partial x} + \frac{\partial}{\partial y}})\, g(x,y) = (e^{D+E})\, g(x,y),$$

where D, E^h, $D^r E^s$ represent respectively

$$\frac{\partial}{\partial x} g(x,y),\quad \frac{\partial^h}{\partial x^h} g(x,y),\quad \frac{\partial^{r+s}}{\partial x^r\, \partial y^s} g(x,y).$$

Hence

$$g(x+1,y+1) - g(x+1,y) - g(x,y+1) + g(x,y)$$
$$= (e^{D+E} - e^{D} - e^{E} + 1)\, g(x,y)$$
$$= \{(e^{D} - 1)(e^{E} - 1)\}\, g(x,y).$$

where the D's are operators operating on the function $g(x,y)$. Therefore

$$g(x,y) = \frac{1}{[(e^{D} - 1)(e^{E} - 1)]}\, f(x,y),$$

where the operators are now operating upon the function $f(x,y)$.

To develop $\frac{1}{(e^u-1)(e^v-1)}$ into a Taylor series it is necessary to develop $\frac{uv}{(e^u-1)(e^v-1)}$ into a Taylor series and then divide by uv. This becomes after $\mathbb{D}$, $\mathbb{E}$ replace u and v respectively,

$$\left\{\frac{1}{[(e^{\mathbb{D}}-1)(e^{\mathbb{E}}-1)]}\right\} f(x,y)=\left\{\frac{1}{\mathbb{D}\mathbb{E}}-\frac{1}{2\mathbb{E}}-\frac{1}{2\mathbb{D}}+\frac{1}{2!}\frac{\mathbb{D}}{6\mathbb{E}}+\frac{1}{2}+\frac{\mathbb{E}}{6\mathbb{D}}\right.$$

$$-\frac{\mathbb{E}^3}{24}-\frac{\mathbb{D}^3}{24}-\frac{\mathbb{D}^3}{720\mathbb{E}}+\frac{\mathbb{D}\mathbb{E}}{144}-\frac{\mathbb{E}^3}{720\mathbb{D}}+\frac{\mathbb{D}^3}{1440}+\frac{\mathbb{E}^3}{1440}$$

$$\left.+\frac{\mathbb{D}^5}{6!42\mathbb{E}}-\frac{\mathbb{D}\mathbb{E}^3}{6!12}-\frac{\mathbb{D}^3\mathbb{E}}{6!12}+\frac{\mathbb{E}^5}{6!42\mathbb{D}}\cdots\right\} f(x,y),$$

where $\frac{1}{\mathbb{D}}$, $\frac{1}{\mathbb{E}}$ represent integration.

Using these results $\sum_c^d \sum_a^b f(x,y)=g(x,y)\Big]_c^{d+1}{}_a^{b+1}$ or

$$\sum_c^d \sum_a^b f(x,y)=\int_c^{d+1}\int_a^{b+1} f(x,y)\,dx\,dy-\frac{1}{2}\int_c^{d+1} f(x,y)\,dy\left[\int_a^{b+1}-\frac{1}{2}\int_a^{b+1} f(x,y)\,dx\right]_c^{d+1}$$

$$+\frac{\partial}{12\,\partial y}\int_a^{b+1} f(x,y)\,dx\Big]_c^{d+1}+\frac{1}{12}\cdot\frac{\partial}{\partial x}\int_c^{d+1} f(x,y)\,dy\Big]_a^{b+1}+\frac{f(x,y)}{4}\Big]_c^{d+1}\Big]_a^{b+1}$$

$$-\frac{\partial f(x,y)}{24\,\partial x}\Big]_c^{d+1}\Big]_a^{b+1}-\frac{\partial f(x,y)}{24\,\partial y}\Big]_c^{d+1}\Big]_a^{b+1}-\frac{\partial^3}{720\,\partial x^3}\int_c^{d+1} f(x,y)\,dy\Big]_a^{b+1}$$

$$+\frac{\partial^2 f(x,y)}{144\,\partial x\,\partial y}\Big]_c^{d+1}\Big]_a^{b+1}-\frac{\partial^3}{720\,\partial y^3}\int_a^{b+1} f(x,y)\,dx\Big]_c^{d+1}$$

$$+\frac{\partial^3 f(x,y)}{1440\,\partial x^3}\Big]_c^{d+1}\Big]_a^{b+1}+\frac{\partial^3 f(x,y)}{1440\,\partial y^3}\Big]_c^{d+1}\Big]_a^{b+1}\cdots\cdots$$

W. D. Baten

THE STANDARD ERROR OF A MULTIPLE REGRESSION EQUATION[1]

By

JOHN RICE MINER

Since a multiple regression equation is essentially a hyperplane, fitted by the method of least squares, its standard error may be obtained from Gauss' *standard error of a function* recently discussed by Schultz (1930). Let the equation be

$$x_1 = b_{12.34\cdots m}x_2 + b_{13.24\cdots m}x_3 + \cdots + b_{1m.23\cdots(m-1)}x_m$$

where x_1 is the dependent variable, x_2, x_3 $\cdots$ x_m the independent variables, each measured from its respective mean, and $b_{12.34\cdots m}, \cdots, b_{1m.23\cdots(m-1)}$ the partial regression coefficients. Then the determinant of Schultz's equation (10) becomes

$$D = \begin{vmatrix} n & 0 & 0 & \cdots & 0 \\ 0 & \Sigma x_2^2 & \Sigma x_2x_3 & \cdots & \Sigma x_2x_m \\ 0 & \Sigma x_2x_3 & \Sigma x_3^2 & \cdots & \Sigma x_3x_m \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & \Sigma x_2x_m & \Sigma x_3x_m & \cdots & \Sigma x_m^2 \end{vmatrix} = n^m\sigma_2^2\sigma_3^2\cdots\sigma_m^2$$

(1)

$$\begin{vmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & r_{23} & \cdots & r_{2m} \\ 0 & r_{23} & 1 & \cdots & r_{3m} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & r_{2m} & r_{3m} & \cdots & 1 \end{vmatrix} = n^m\sigma_2^2\sigma_3^2\cdots\sigma_m^2\Delta, \text{sau}$$

[1]From the Department of Biology of the School of Hygiene and Public Health of the Johns Hopkins University.

Let Δ_{ij} be the cofactor of the element in the i'th row and the j'th column. Then

$$[\alpha\alpha] = D_{11}/D = \frac{1}{n},$$

$$[\beta\beta] = D_{22}/D = \Delta_{22}/n\sigma_2^2\Delta,$$

. . . . . .

$$[\mu\mu] = D_{mm}/D = \Delta_{mm}/n\sigma_m^2\Delta,$$

$$[\alpha\beta] = D_{12}/D = 0,$$

. . . . . .

$$[\alpha\mu] = D_{1m}/D = 0,$$

$$[\beta\gamma] = D_{23}/D = \Delta_{23}/n\sigma_2\sigma_3\Delta,$$

. . . . . .

$$[\beta\mu] = D_{2m}/D = \Delta_{2m}/n\sigma_2\sigma_m\Delta,$$

. . . . . .

$$\frac{\partial f}{\partial A} = 1, \frac{\partial f}{\partial B} = x_2, \frac{\partial f}{\partial C} = x_3, \cdots, \frac{\partial f}{\partial M} = x_m, \quad \text{and}$$

$$\varepsilon^2 = \frac{n}{n-m}\sigma_1^2(1 - R_{1(23\cdots m)}^2).$$

Therefore, substituting these values in Schultz's equation (27.1), we have

$$\sigma_f = \frac{\sigma_1}{(n-m)^{\frac{1}{2}}}\left(1-R^2_{1(23\ldots m)}\right)^{\frac{1}{2}}\left\{1+\frac{\Delta_{22}}{\sigma_2^2\Delta}x_2^2+\frac{\Delta_{33}}{\sigma_3^2\Delta}x_3^2+\ldots+\frac{\Delta_{mm}}{\sigma_m^2\Delta}x_m^2\right.$$

(2)

$$\left.+2\frac{\Delta_{23}}{\sigma_2\sigma_3\Delta}x_2x_3+\ldots+2\frac{\Delta_{2m}}{\sigma_2\sigma_m\Delta}x_2x_m+\ldots\right\}^{\frac{1}{2}}.$$

For a simple regression equation this reduces to

$$(3)\quad \sigma_f = \frac{\sigma_1}{(n-2)^{\frac{1}{2}}}\left(1-r_{12}^2\right)^{\frac{1}{2}}\left\{1+\frac{x_2^2}{\sigma_2^2}\right\}^{\frac{1}{2}}.$$

This agrees with the expression given by Pearson (1913), if we remember that x_2 is measured from its mean and that Pearson does not correct for the number of parameters.

For a regression equation with two independent variables

$$\sigma_f = \frac{\sigma_1}{(n-3)^{\frac{1}{2}}}\left(1-R^2_{1(23)}\right)^{\frac{1}{2}}$$

$$(4)\quad \left\{1+\frac{x_2^2}{\sigma_2^2(1-r_{23}^2)}+\frac{x_3^2}{\sigma_3^2(1-r_{23}^2)}-\frac{2r_{23}x_2x_3}{\sigma_2\sigma_3(1-r_{23}^2)}\right\}^{\frac{1}{2}}$$

$$=\frac{\sigma_{1.23}}{(n-3)^{\frac{1}{2}}}\left\{1+\frac{x_2^2}{\sigma_{2.3}^2}+\frac{x_3^2}{\sigma_{3.2}^2}-\frac{2r_{23}x_2x_3}{\sigma_{2.3}\sigma_{3.2}}\right\}^{\frac{1}{2}}.$$

As an example of the application of this formula we may calculate the standard error of the mean heart-weight (X_1) of the array of persons with an age (X_2) of 52.92 years and a

body-weight (X_3) of 49.93 kilograms in a population of 213 persons characterized by the following biometric constants:

$M_1 = 348.9$ g; $\sigma_1 = 79.4$ g; $r_{12} = +0.114$

$M_2 = 59.65$ yrs.; $\sigma_2 = 17.54$ yrs.; $r_{13} = +0.652$

$M_3 = 56.45$ kg; $\sigma_3 = 14.38$ kg; $r_{23} = -0.185$.

From these data $r_{12.3} = +0.315$ and $r_{13.2} = +0.689$ and the regression equation of heart-weight on age and body-weight is

$$X_1 = 66.09 + 1.100X_2 + 3.848X_3$$

from which the mean heart-weight of persons aged 52.92 years and weighing 49.93 kg. is 316.4 g.

Substituting the appropriate values of the constants in (4) and remembering that $x_2 = X_2 - M_2 = -6.72$, $x_3 = X_3 - M_3 = -6.52$, and

$$(1-R^2_{1(23)})^{\frac{1}{2}} = (1-r^2_{12})^{\frac{1}{2}}(1-r^2_{13.2})^{\frac{1}{2}}$$

$$\sigma_f = \frac{79.4}{210^{\frac{1}{2}}}\ (0.993)(0.725)\left\{1+\frac{(-6.72)^2}{(17.54)^2(0.966)}+\frac{(-6.52)^2}{(14.38)^2(0.966)}\right.$$

$$\left.-\frac{2(-0.185)(-6.72)(-6.52)}{(17.54)(14.38)(0.966)}\right\}^{\frac{1}{2}} = 4.7\ g.$$

John Rice Miner

REFERENCES

Pearson, Karl. 1913. On the probable errors of frequency constants. Biometrika, 9:1-10.

Schultz, Henry. 1930. The standard error of a forecast from a curve. J. Amer. Stat. Assoc., 25:139-185.

SAMPLING IN THE CASE OF CORRELATED OBSERVATIONS

By

CECIL C. CRAIG

National Research Fellow

Dr. E. C. Rhodes, in a paper in the *Journal of the Royal Statistical Society*,[1] has considered the distribution of characteristics of samples of N when the individual observations are not assumed to be independent. As he points out, there are many important cases in which the usual assumption of independence or randomness in the observations is not justifiable. In the present paper will be explained a method based on the semi-invariants of Thiele for the calculation of the characteristics of the sought distributions in this case which is especially to be preferred to the method based on moments when it is supposed that the observations are normally correlated. In the case it is further assumed that only consecutive observations are correlated, in addition to Dr. Rhodes' results, the third semi-invariant (which is the same as the third moment about the mean) of the variance and the mean and the variance of the third and fourth moments about the mean are given.

Let the N observations composing a sample be given by values of $x_1, x_2, \cdots, x_N$ respectively, and let $F_N(x_1, x_2, \cdots, x_N)$ be the n-way probability function of $x_1, x_2, \cdots$, and x_N.

[1]The Precision of Means and Standard Deviations When the Individual Errors Are Correlated, Vol. 90 (1927), pp. 135-143.

Then the semi-invariants, $\lambda_{rst\cdots}$ of $x_1, x_2, \cdots, x_N$ are defined by

$$e^{\left(\sum_1^N \lambda_i t_i\right)+\frac{1}{2}\left(\sum_1^N \lambda_i t_i\right)^{(2)}+\frac{1}{3!}\left(\sum_1^N \lambda_i t_i\right)^{(3)}+\cdots}$$

$$(1) \qquad = \int_{-\infty,\cdots,-\infty}^{\infty,\cdots,\infty} dF_N(x_1, x_2, \cdots, x_N)\, e^{\left(\sum_1^N x_i t_i\right)} \qquad [1]$$

which is to be regarded as a formal identity in $t_1, t_2, \cdots, t_N$. $\left(\sum_1^N \lambda_i t_i\right)^{(\kappa)}$ is first expanded by the multinomial law and then each term $\lambda_1^r, \lambda_2^s, \lambda_3^t \cdots$ in the result is replaced by $\lambda_{rst\cdots}$.

We shall pass over the characteristics of distributions of means, since the method of semi-invariants is equivalent to that of moments in this case, and take up the distribution of moments about the mean in samples of N. Following the method previously used by the author in the case of independent observations,[2] let

$$\delta_i = x_i - \sum_1^N \frac{x_i}{N}$$

$$(2) \qquad = \sum_{j=1}^N a_{ij} x_j \quad \text{with} \quad \begin{cases} a_{ij} = -\frac{1}{N} \\ a_{ii} = \frac{N-1}{N} \end{cases}.$$

Then let $V(\delta_1, \delta_2, \cdots, \delta_{N-1})$ be the probability function of $\delta_1, \delta_2, \cdots, \delta_{N-1}, (\sum \delta_i = 0$. The semi-invariants $\lambda'_{rst\cdots}$ of $\delta_1, \delta_2, \cdots, \delta_{N-1}$ are defined by

---

[1] Following Cramér, I distinguish between probability and frequency functions. $F_N(x_1, x_2, \cdots, x_N)$ is the "cumulative" frequency function and thus the integral is an n-way Stieltjes integral.

[2] An Application of Thiele's Semi-invariants to the Sampling Problem; Metron, Vol. 7, No. 4 (1928), pp. 3-75.

$$e^{\left(\sum_1^{N-1}\lambda_i' t_i\right)+\frac{1}{2}\left(\sum_1^{N-1}\lambda_i' t_i\right)^{(2)}+\frac{1}{3!}\left(\sum_1^{N-1}\lambda_i' t_i\right)^{(3)}+\cdots}$$

$$(3)\qquad =\int_{-\infty,\cdots,-\infty}^{\infty,\cdots,\infty} dV(\delta_1,\cdots,\delta_{N-1})\,e^{\left(\sum_1^{N-1}\delta_i t_i\right)}$$

$$=\int_{-\infty,\cdots,-\infty}^{\infty,\cdots,\infty} dF_N(x_1,x_2,\cdots,x_N)\,e^{\left(\sum_{i=1}^{N-1}\sum_{j=1}^{N}a_{ij}x_j t_i\right)}$$

We have at once,

$$\left(\sum_{i=1}^{N-1} t_i \sum_{j=1}^{N}\lambda_j a_{ij}\right)^{(K)}=\left(\sum_{i=1}^{N-1}\lambda_i' t_i\right)^{(K)}$$

and as the author has previously remarked,[1] we can also write

$$(4)\qquad \left(\sum_1^N t_i \sum_{j=1}^{N}\lambda_j a_{ij}\right)^{(K)}=\left(\sum_1^N \lambda_i' t_i\right)^{(K)}$$

so long as the relation is only used to find the values of $\lambda'_{rst\cdots}$'s in which at least one of the subscripts is zero.

Then $S_K(\nu_n)$, the k'th semi-invariant of the n'th moment about the mean in samples of N, is given by the formula

$$S_K(\nu_n)=$$

$$\frac{1}{N^K}\Sigma\Sigma\cdots\frac{(-1)^{(r+s+t+\cdots)-1}\left[(r+s+t+\cdots)-1\right]!K!\,\nu'^{r}_{a_1n_1,a_2n_1,\cdots}\nu'^{s}_{b_1n_1,b_2n_1,\cdots}\nu'^{t}_{c_1n_1c_2n_1,\cdots}\cdots}{\left[a_1!\,a_2!\cdots\right]^r\left[b_1!\,b_2!\cdots\right]^s\left[c_1!\,c_2!\cdots\right]^t\,r!s!t!\cdots}$$

[1] loc. cit., pp. 18, 19.

the notation $\nu'_{u\,v\,w}$ referring to moments of $\delta_1, \cdots \delta_{N-1}, \delta_N$, the summation including all terms for which

$$r(a_1 + a_2 + \cdots) + s(b_1 + b_2 + \cdots) + t(c_1 + c_2 + \cdots) + \cdots = k,$$

$$a_1 \geq a_2 \geq \cdots$$

$$b_1 \geq b_2 \geq \cdots$$

$$c_1 \geq c_2 \geq \cdots$$

$$\cdots\cdots$$

$$(a_1 + a_2 + \cdots) \geq (b_1 + b_2 + \cdots) \geq (c_1 + c_2 + \cdots) \geq \cdots$$

In particular:

$$S_1(\nu_n) = \frac{1}{N} \Sigma \nu'_{n,0}, \quad (\Sigma \nu'_{n,0} = \nu'_{n,0\cdots 0} + \nu'_{0,n,0\cdots 0} + \cdots + \nu'_{q,q,n,0\cdots 0} + \cdots),$$

$$S_2(\nu_n) = \frac{1}{N^2}\left[\Sigma \nu'_{2,0} + 2\Sigma \nu'_{n,n,0} - (\Sigma \nu'_{n,0})^2\right],$$

$$S_3(\nu_n) = \frac{1}{N^3}\left[\Sigma \nu'_{3n,0} + 3\Sigma \nu'_{2n,n,0} + 6\Sigma \nu'_{n,n,n,0} - 3(\Sigma \nu'_{2n,0})(\Sigma \nu'_{n,0}) - 6(\Sigma \nu'_{n,n,0})(\Sigma \nu'_{n,0}) + 2(\Sigma \nu'_{n,0})^3\right],$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

On writing out the moments $\nu'_{u\,v\,w}$ in terms of the semi-invariants $\lambda'_{r\,s\,t}$[2] and then using (4) the sought semi-invariants are obtained.

In the case that the N observations are normally correlated and $F_N(x_1, x_2, \cdots, x_N)$ is the N-dimensional normal probability function, the left-hand member of (4) vanishes for $k \geq 3$.

If we suppose that the standard deviations of $x_1, x_2, \cdots, x_N$ are all equal (which we shall always do) and take as the simplest case that $x_1, x_2, \cdots, x_N$ are normally correlated and that

[1] See the author's paper cited, p. 21, formula (25).

[2] For a detailed explanation of this kind of calculation see the author's paper cited, pp. 23-27.

the correlation as measured by the Pearsonian coefficient, $r_{x_i x_j}$, is the same for each pair, x_i, x_j, of the set of N observations, we get

$$\lambda'_{20} = \lambda'_{020} = \lambda'_{0020} = \cdots = \frac{N-1}{N}(\lambda_{20} - \lambda_{11}) = \frac{N-1}{N}(1-r)\lambda_{20},$$

$$\lambda'_{110} = \lambda'_{1010} = \lambda'_{0110} = \cdots = -\frac{1}{N}(\lambda_{20} - \lambda_{11}) = -\frac{1}{N}(1-r)\lambda_{20},$$

if the common value of $r_{x_i x_j}$ be denoted simply by r. But if the observations are independent and the parent population is normal we have

$$\lambda'_{20} = \lambda'_{020} = \lambda'_{0020} = \cdots = \frac{N-1}{N}\lambda_{20},$$

$$\lambda'_{110} = \lambda'_{1010} = \lambda'_{0110} = \cdots = -\frac{1}{N}\lambda_{20}.$$

Thus it follows that the distributions of the characteristics of samples of N in this particular case of dependent observations are the same as if the observations were independent and taken from a normal population of variance $(1-r)\lambda_{20}$.

In case $F_N(x_1, x_2, \cdots, x_N)$ is normal it is convenient to express the right hand members of (5) directly in terms of the semi-invariants $\lambda'_{rst\ldots}$ for $n = 2, 3, 4$. For that purpose we shall adopt the following notation. Let the linear form $\sum_{j=1}^{N} a_{ij}\lambda_j$ be denoted by A_i. Then (4) becomes

$$(6) \qquad \left(\sum_1^N A_i t_i\right)^{(\kappa)} = \left(\sum_1^N \lambda'_i t_i\right)^{(\kappa)}.$$

Thus in a symbolic sense A_i's and λ_i's are equivalent. But with regard to the subscripts of the A terms in the expansion of the left member of (6) we use a different convention than for the subscripts of the λ's. We set

[1] See the author, loc. cit., p. 19.

$$\lambda'_{20} = A_{11}\ ,\ \lambda_{020} = A_{22}, \ldots\ ;$$

$$\lambda'_{110} = A_{12}\ ,\ \lambda_{1010} = A_{13}, \ldots\ .$$

We get

$$S_1(\nu_2) = \frac{1}{N}\Sigma A_{ii}\ ,$$

$$S_2(\nu_2) = \frac{1}{N^2}\left[3\Sigma A_{ii}^2 + 2\Sigma A_{ii}A_{jj} + 4\Sigma A_{ij}^2 - (\Sigma A_{ii})^2\right], i \neq j,$$

the summations, of course, running over all values of i and j from 1 to N. But since

$$\Sigma A_{ii}^2 + 2\Sigma A_{ii}A_{jj} = (\Sigma A_{ii})^2$$

the second relation reduces to

$$S_2(\nu_2) = \frac{2}{N^2}\left(\Sigma A_{ii}^2 + 2\Sigma A_{ij}^2\right).$$

Similarly

$$S_3(\nu_2) = \frac{8}{N^3}\left(\Sigma A_{ii}^3 + 3\Sigma A_{ii}A_{ij}^2 + 6\Sigma A_{ij}A_{ik}A_{jk}\right),$$

$$S_4(\nu_2) = \frac{48}{N^4}\left(\Sigma A_{ii}^4 + 4\Sigma A_{ii}^2A_{ij}^2 + 4\Sigma A_{ii}A_{jj}A_{ij}^2 + 2\Sigma A_{ij}^4\right.$$

$$\left. + 8\Sigma A_{ii}A_{ij}A_{ik}A_{jk} + 4\Sigma A_{ij}^2A_{ik}^2 + 8\Sigma A_{ij}A_{ik}A_{jl}A_{kl}\right), \tag{7}$$

. . . . . .

$$S_1(\nu_3) = 0,$$

$$S_2(\nu_3) = \frac{3}{N^2}\left(5\Sigma A_{ii}^3 + 6\Sigma A_{ii}A_{jj}A_{ij} + 4\Sigma A_{ij}^3\right),$$

$$S_3(\nu_3) = 0,$$

. . . . . .

$$S_1(V_4) = \frac{3}{N}\Sigma A_{ii}^2 ,$$

$$S_2(V_4) = \frac{48}{N^2}\left(2\Sigma A_{ii}^4 + 3\Sigma A_{ii} A_{jj} A_{ij}^2 + \Sigma A_{ij}^2\right).$$

To illustrate the use of these formulas and to give some results in a case of practical interest, let us suppose that the set of N observations composing a sample may be assigned an order in which only consecutive observations are correlated and in a constant degree. Thus our observations might be prices or indices taken at the ends of consecutive time intervals. We suppose, then, that

$$\lambda_{110} = \lambda_{0110} = \lambda_{00110} = \cdots = r\lambda_{20} ,$$

$$\lambda_{101} = \lambda_{1001} = \lambda_{0101} = \cdots = 0 .$$

The first step in the calculation is to obtain the values of the various A's which enter into the formulas (7). A_{11} is found from A_1^2, A_{12} from $A_1 A_2$ and so on. We get

$$\begin{aligned}
&A_{11} = A_{N,N} = \left(1 - \frac{1}{N} - \frac{2r}{N^2}\right)\lambda_{20} ,\\
&A_{22} = A_{33} = \cdots = A_{N-1,N-1} = \left(1 - \frac{1+2r}{N} - \frac{2r}{N^2}\right)\lambda_{20} ,\\
&A_{12} = A_{N-1,N} = \left(r - \frac{1+r}{N} - \frac{2r}{N^2}\right)\lambda_{20} ,\\
&A_{23} = A_{34} = \cdots = A_{N-2,N-1} = \left(r - \frac{1+2r}{N} - \frac{2r}{N^2}\right)\lambda_{20} ,\\
&A_{13} = A_{14} = \cdots = A_{1,N-1} ,\\
&\quad = A_{2,N} = A_{3,N} = \cdots = A_{N-2,N} = \left(-\frac{1+r}{N} - \frac{2r}{N^2}\right)\lambda_{20} ,\\
&A_{1,N} = \left(-\frac{1}{N} - \frac{2r}{N^2}\right)\lambda_{20} ,\\
&A_{ij} = \left(-\frac{1+2r}{N} - \frac{2r}{N^2}\right)\lambda_{20} \qquad \begin{cases} 1 < i < N-1 \\ 1 < j < N-1 \\ |i-j| > 1 \end{cases}
\end{aligned} \tag{8}$$

Then, on substitution in (7), we have finally

$$S_1(V_2) = (1-\tfrac{1}{N})(1-\tfrac{2r}{N})\lambda_{20},$$

$$S_2(V_2) = \tfrac{2}{N}\left[(1-\tfrac{1}{N})(1-\tfrac{4r}{N})+2r^2(1-\tfrac{3}{N}+\tfrac{2}{N^2}+\tfrac{2}{N^3})\right]\lambda_{20}^2.$$

These two results are given by Dr. Rhodes, loc. cit., though there is a slight misprint in the second one as given there. The remainder of the results given here are believed to be new.

$$S_3(V_2) = \tfrac{8}{N^2}\left[(1-\tfrac{1}{N})(1-\tfrac{6r}{N})+6r^2(1-\tfrac{3}{N}+\tfrac{2}{N^2}+\tfrac{2}{N^3})-\tfrac{4r^3}{N}(2-\tfrac{3}{N}-\tfrac{3}{N^2}-\tfrac{2}{N^3})\right]\lambda_{20}^3.$$

$$S_1(V_3) = 0,$$

$$S_2(V_3) = \tfrac{6}{N}\left[(1-\tfrac{1}{N})(1-\tfrac{2}{N})(1-\tfrac{6r}{N})-\tfrac{6r^2}{N}(1-\tfrac{5}{N}+\tfrac{12}{N^3})+\tfrac{2r^3}{N^2}(1-\tfrac{7}{N}+\tfrac{14}{N^2}\right.$$

$$\left.+\tfrac{2}{N^3}-\tfrac{24}{N^4}-\tfrac{40}{N^5})\right]\lambda_{20}^4,$$

$$S_3(V_3) = 0,$$

$$S_1(V_4) = 3\left[(1-\tfrac{1}{N})^2(1-\tfrac{4r}{N})+\tfrac{4r^2}{N^2}(1-\tfrac{3}{N^2})\right]\lambda_{20}^2,$$

$$S_2(V_4) = \tfrac{24}{N}\left[(1-\tfrac{1}{N})(4-\tfrac{9}{N}+\tfrac{6}{N^2})(1-\tfrac{8r}{N})+6r^2(1-\tfrac{5}{N}+\tfrac{25}{N^2}-\tfrac{29}{N^3}-\tfrac{44}{N^4}+\tfrac{68}{N^5})\right.$$

$$-\tfrac{8r^3}{N}(4-\tfrac{19}{N}+\tfrac{33}{N^2}+\tfrac{30}{N^3}-\tfrac{54}{N^4}-\tfrac{108}{N^5})+2r^4(1-\tfrac{9}{N}+\tfrac{44}{N^2}-\tfrac{64}{N^3}$$

$$\left.-\tfrac{114}{N^4}+\tfrac{192}{N^5}+\tfrac{360}{N^6}+\tfrac{288}{N^7})\right]\lambda_{20}^4.$$

It should be observed that the expressions for $S_1(\nu_n)$ for $N < 3$ and for $S_k(\nu_n)$, $k \geq 2$ for $N < 5$ are in general not valid, since it can be seen by reference to (8) that all the types of A's used in the formulas (7) do not exist for values of N so small. But for these small values of N, the values of the characteristics for which expressions are given above can be readily computed directly.

C. C. Craig

Stanford University.

THE RELATION BETWEEN THE MEANS AND VARIANCES, MEANS SQUARED AND VARIANCES IN SAMPLES FROM COMBINATIONS OF NORMAL POPULATIONS

By

G. A. BAKER

The distributions of the means and variances of samples from the combinations of normal populations have been discussed in a previous paper.[1] It is known that if the sampled population is not normal the means and variances of samples are not independent.

The present discussion aims to give some idea of the relation between the means and the variances, means squared and variances of samples from a population that is the combination of normal populations. To this end the case of samples of two from such populations is rather completely investigated. Also empirical random sampling results for two special populations are presented.

Suppose that a population is represented by

$$(1) \qquad f(x)=\frac{1}{1+k}\left[\frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}}+\frac{k}{\sigma\sqrt{2\pi}}\,e^{-\frac{1}{2}\frac{(x-m)^2}{\sigma^2}}\right].$$

[1] "Random Sampling from Non-Homogeneous Populations," Metron, Vol. VIII, No. 3 (1930), pp. 1-21.

If a method used by Karl Pearson[1] is followed, the probability of

$$x_1 \text{ in } dx_1 \text{ is } f(x_1)dx_1 ,$$

$$x_2 \text{ in } dx_2 \text{ is } f(x_2)dx_2$$

and the probability of the concurrence of these two events is

$$f(x_1)\,f(x_2)dx_1\,dx_2 \tag{2}$$

which may be written

$$\frac{1}{(1+k)^2 2\pi}\left[e^{-\frac{1}{2}\left[x_1^2+x_2^2\right]}+\frac{k^2}{\sigma^2}\,e^{-\frac{1}{2\sigma^2}\left[(x_1-m)^2+(x_2-m)^2\right]}\right.$$

(3)

$$\left.+\frac{k}{\sigma}\left\{e^{-\frac{1}{2}\left[x_1^2+\frac{(x_2-m)^2}{\sigma^2}\right]}+e^{-\frac{1}{2}\left[x_2^2+\frac{(x_1-m)^2}{\sigma^2}\right]}\right\}\right]dx_1dx_2$$

Now

$$x=\tfrac{1}{2}(x_1+x_2)$$

$$\Sigma^2=\tfrac{1}{2}\left[(x_1-x)^2+(x_2-x)^2\right] .$$

Whence

$$\begin{cases} x_1=-\Sigma+x \\ x_2=\Sigma+x \end{cases} \tag{4}$$

Also $dx_1\,dx_2$ may be replaced[2] by

[1]Appendix to Papers by "Student" and R. A. Fisher, Biometrika, Vol. XIX (1925), p. 522.

[2]R. A. Fisher: "Frequency Distribution of the Values of the Correlation Coefficient in Samples from an Indefinitely Large Population," Biometrika, Vol. X (1915), p. 507.

(5) $$c\, d\Sigma\, dx$$

In virtue of (5), (4) and (3), (6) is obtained.

(6)
$$\frac{1}{(1+k)^2 2\pi}\left[e^{-\frac{1}{2}\left[2x^2+2\Sigma^2\right]}+\frac{k^2}{\sigma^2}e^{-\frac{1}{2\sigma^2}\left[2\Sigma^2+2(x-m)^2\right]}\right.$$

$$+\frac{k}{\sigma}\left\{e^{-\frac{1}{2}\left[(-\Sigma+x)^2+\frac{(\Sigma+x-m)^2}{\sigma^2}\right]}\right.$$

$$\left.\left.+e^{-\frac{1}{2}\left[(\Sigma+x)^2+\frac{(-\Sigma+x-m)^2}{\sigma^2}\right]}\right\}\right].$$

This is the correlation surface for the means and standard deviations of samples of two drawn from (1). To get the correlation surface of the means and variances write

$$\Sigma^2 = u$$

$$d\Sigma = \frac{du}{2\sqrt{u}}$$

Then

(7)
$$F(x,u)=\frac{1}{(1+k)^2 2\pi}\left[\frac{e^{-\frac{1}{2}\left[2x^2+2u\right]}}{2\sqrt{u}}+\frac{k^2}{2\sqrt{u}\,\sigma^2}e^{-\frac{1}{2\sigma^2}\left[2u+2(x-m)^2\right]}\right.$$

$$+\frac{k}{\sigma}\left\{\frac{e^{-\frac{1}{2}\left[(\sqrt{u}+x)^2+\frac{(-\sqrt{u}+x-m)^2}{\sigma^2}\right]}}{2\sqrt{u}}\right.$$

$$\left.\left.+\frac{e^{-\frac{1}{2}\left[(-\sqrt{u}+x)^2+\frac{(\sqrt{u}+x-m)^2}{\sigma^2}\right]}}{2\sqrt{u}}\right\}\right]$$

is the desired surface.

The locus of mean u's for given x's is

$$u = \frac{e^{-x^2+k^2\sigma e^{\frac{-(x-m)^2}{\sigma^2}+\frac{4\sqrt{2}k}{(\sigma^2+1)^{\frac{3}{2}}}}} e^{\frac{-2(x-\frac{m}{2})^2}{\sigma^2+1}}\left[\sigma^2+\frac{\{(\sigma^2-1)x+m\}^2}{\sigma^2+1}\right]}{e^{-x^2+\frac{k^2}{\sigma}} e^{\frac{-(x-m)^2}{\sigma^2}+\frac{2\sqrt{2}k}{\sqrt{\sigma^2+1}}} e^{\frac{-2(x-\frac{m}{2})^2}{\sigma^2+1}}}. \tag{8}$$

The locus of the mean x's for given u's is

$$x = \frac{\frac{mk}{\sigma}e^{-\frac{u}{\sigma^2}+\frac{2\sqrt{2}k}{\sqrt{\sigma^2+1}}}\left[\left\{(\sigma^2-1)\sqrt{u}+m\right\}e^{\frac{-2(\sqrt{u}-\frac{m}{2})^2}{\sigma^2+1}}-\left\{(\sigma^2-1)\sqrt{u}-m\right\}e^{\frac{-2(\sqrt{u}+\frac{m}{2})^2}{\sigma^2+1}}\right]}{e^{-u+\frac{k^2}{\sigma}}e^{-\frac{u}{\sigma^2}+\frac{2\sqrt{2}k}{\sqrt{\sigma^2+1}}}\left\{e^{\frac{-2(\sqrt{u}-\frac{m}{2})^2}{\sigma^2+1}}+e^{\frac{-2(-\sqrt{u}-\frac{m}{2})^2}{\sigma^2+1}}\right\}}. \tag{9}$$

The correlation surface for the means squared (= z) and variances is

$$\psi(u,z) = \frac{1}{(1+k)^2 2\pi}\left[\frac{e^{-\frac{1}{2}[2z+2u]}}{4\sqrt{u}\sqrt{z}}+\frac{k^2 e^{\frac{-1}{2\sigma^2}[2u+2(\sqrt{z}-m)^2]}}{\sigma^2 4\sqrt{u}\sqrt{z}}\right.$$
$$+\frac{k}{\sigma}\left\{\frac{e^{-\frac{1}{2}[(\sqrt{u}+\sqrt{z})^2+\frac{(-\sqrt{u}+\sqrt{z}-m)^2}{\sigma^2}]}}{4\sqrt{u}\sqrt{z}}\right. \tag{10}$$
$$\left.\left.+\frac{e^{-\frac{1}{2}[(-\sqrt{u}+\sqrt{z})^2+\frac{(\sqrt{u}+\sqrt{z}-m)^2}{\sigma^2}]}}{4\sqrt{u}\sqrt{z}}\right\}\right].$$

The locus of the mean u's for given z's is

$$(11)\quad u = \frac{e^{-z+k^2}\sigma e^{\frac{-(\sqrt{2}z-m)^2}{\sigma^2}} + \frac{4\sqrt{2}k}{\sqrt{(\sigma^2+1)^3}} e^{\frac{-2(\sqrt{2}z-\frac{m}{2})^2}{\sigma^2+1}\left[\sigma^2+\frac{\{(\sigma^2-1)\sqrt{2}z+m\}^2}{\sigma^2+1}\right]}}{e^{-z+\frac{k^2}{\sigma}} e^{\frac{-(\sqrt{2}z-m)^2}{\sigma^2}} + \frac{2\sqrt{2}k}{\sqrt{\sigma^2+1}} e^{\frac{-2(\sqrt{2}z-\frac{m}{2})^2}{\sigma^2+1}}}$$

The locus of the mean z's for given u's is

$$z = \frac{1}{e^{-u+\frac{k^2}{\sigma}} e^{-\frac{u}{\sigma^2}} + \frac{2\sqrt{2}k}{\sqrt{\sigma^2+1}}\left\{e^{\frac{-2(\sqrt{u}-\frac{m}{2})^2}{\sigma^2-1}} + e^{\frac{-2(\sqrt{u}+\frac{m}{2})^2}{\sigma^2+1}}\right\}}$$

(12) multiplied by

$$\left[e^{-u+\frac{k^2}{\sigma}(\sigma^2+m^2)} e^{-\frac{u}{\sigma^2}} + \frac{4\sqrt{2}k}{(\sigma^2+1)^{\frac{3}{2}}}\left\{\left(\sigma^2+\frac{[(\sigma^2-1)\sqrt{n}-m]^2}{\sigma^2+1}\right) e^{\frac{-2(\sqrt{u}-\frac{m}{2})^2}{\sigma^2+1}} + \left(\sigma^2+\frac{[(\sigma^2-1)\sqrt{u}+m]^2}{\sigma^2+1}\right) e^{\frac{-2(\sqrt{u}+\frac{m}{2})^2}{\sigma^2+1}}\right\}\right]$$

By expanding the denominators of (8), (9), (11), and (12) by the multinomial theorem, it can be shown that each of these loci is essentially parabolic, $\sigma^2 - 1 \neq 0$. They are subject to an exponential influence at the beginning of the range of the independent variable, which influence rapidly diminishes as the independent variable takes on higher values.

The probability relations in general between means and variances, means squared and variances will be expected to approximate those for the case of samples of two, because of the fol-

lowing considerations. Suppose that n (the number in the sample) is large.[1] When a large proportion of the sample comes from the first component, the first term of (7) with 2 in the numerator of the exponent replaced by n and with $u^{-\frac{1}{2}}$ replaced by $u^{\frac{n-3}{2}}$ will be an approximation to the surface of the means and variances. Similarly, when a large proportion of the sample comes from the second component, the second term of (7) with 2 in the numerator of the exponent replaced by n and with $u^{-\frac{1}{2}}$ replaced by $u^{\frac{n-3}{2}}$ will be an approximation to the surface of the means and variances. When about equal proportions of the sample come from each component, the last term of (7) with $\frac{2}{2}$ in the numerator of each exponent replaced by $\frac{n}{2}$ and with $u^{-\frac{1}{2}}$ replaced by $u^{\frac{n-3}{2}}$ will be an approximation to the surface of the means and variances. Or, all together, (7) with the mentioned changes in the exponents of the terms, with proper weighting of the terms, and with $u^{-\frac{1}{2}}$ replaced by $u^{\frac{n-3}{2}}$ is a proportionate approximation to the distribution of the means and variances of samples drawn from a population represented by (1). Further, increasing n will not influence relations (8), (9), (11), and (12) as approximations for the general case except the exponential term, if it is assumed that the denominators are expanded and then multiplied by the numerators, for $\frac{1}{n}$ occurs to the same power in the numerators and denominators.

---

[1]Note: The effect of k and of the binomial coefficients is roughly as follows. If the $n+1$ terms denoting s from the first component of (1) and $n-s$ from the second component are divided into thirds, then, if l_1, l_2, l_3 are the exponents of k in the middle terms,

$l_1 = \frac{n-2}{6}$, $l_2 = \frac{3n}{6}$, $l_3 = \frac{3n+2}{6}$

or approximately, since n is large and since only a proportionate expression is desired $l_1' = 0$, $l_2' = \frac{n}{3}$, $l_3 = \frac{2n}{3}$

or the exponents of k of the middle terms of the three sections above are $\frac{n}{3}$ times the exponents of k in (7). The effect of increasing n because of the binomial coefficients is to weight the middle section of the possible surfaces to a much greater extent than the extreme sections, so that with n very large the last term of (7) with 2 replaced by n becomes an approximation to the desired surface.

From (8), (9), (11), and (12) it is clear that the parameters of the sampled population have great influence on the regression relations considered. It should be borne in mind in this connection that many flattened and skewed, as well as bimodal, distributions can be adequately represented by combinations of normal populations. Also, results (8), (9), (11), and (12) can be extended to the sums and differences of any number of normal curves, subject to the condition that the resultant is always positive.

In 1925, Dr. Neyman[1] gave the correlation coefficient between the deviations of the means of samples from the mean of the sampled population and the variances of these samples for samples of n drawn at random from an infinite uni-variate population in terms of the betas of the sampled population as

$$(13) \qquad \rho' = \frac{\sqrt{n-1}\,\sqrt{\beta_1}}{\sqrt{(n-1)\beta_2 - n + 3}} .$$

Similarly, the correlation coefficient between the deviations squared of the means of samples from the mean of the sampled population and the variances is

$$(14) \qquad R' = \frac{\sqrt{n-1}\,(\beta_2 - 3)}{\sqrt{(\beta_2 + 2n - 3)\left[(n-1)\beta_2 - n + 3\right]}} .$$

Under certain very special conditions the statement of ρ' and R' may give an adequate idea of the regression relation between the means and variances, means squared and variances of samples from a population represented by (1). In general the mere statement of these coefficients will not give any useful

[1] J. Splawa-Neyman: "Contributions to the Theory of Small Samples Drawn from a Finite Population," Biometrika, Vol. XVII (1925), pp. 472-479.

notion of the actual probability relations. This is true because: (a) the regression relations between means and variances, means squared and variances of samples from a population represented by (1) are essentially parabolic, as shown for samples of two and as seems probable for larger samples; (b) the frequency arrays may vary markedly in dispersion, in skewness, and in other characteristics.

To illustrate these remarks, samples of four were drawn from two special populations by throwing dice.

Suppose that a population is represented by

$$(15)\quad f(x)=\frac{1}{1+k}\left[\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x+m_1)^2}+\frac{k}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\frac{(x-m_2)^2}{\sigma^2}}\right].$$

The first four moments of $f(x)$ about its mean are

$$\mu_0=1,$$

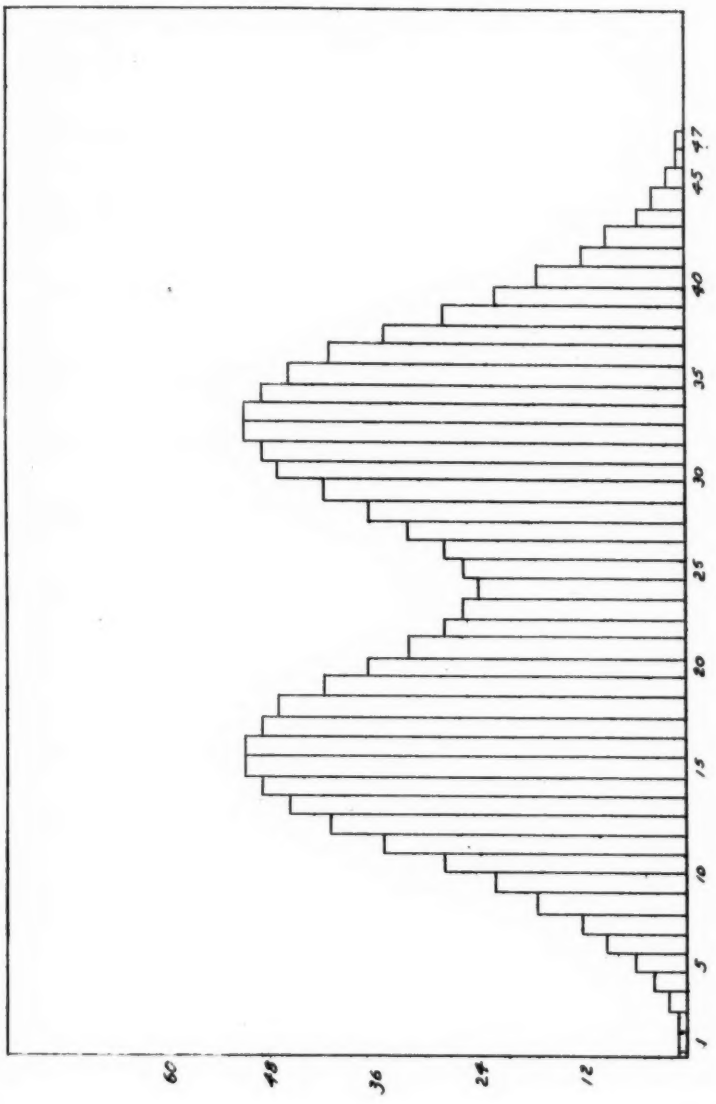
$$\mu_1=\frac{(-m_1+km_2)}{1+k}=0,$$

$$\mu_2=\frac{[1+m_1^2+k(\sigma^2+m_2^2)]}{1+k},$$

$$\mu_3=\frac{-3m_1-m_1^3+k(3m_2\sigma^2+m_2^3)}{1+k},$$

$$\mu_4=\frac{3+6m_1^2+m_1^4+k(3\sigma^4+6m_2^2\sigma^2+m_2^4)}{1+k}.$$

CHART A



Population I, from Which the 1038 Samples of Four of Tables I and II were drawn.

TABLE I

Correlation Table Showing the Relation Between the Means and Variances of Samples of Four from Population I

Means of Samples	*Variances of Samples*																	
	0 to 15	15 to 30	30 to 45	45 to 60	60 to 75	75 to 90	90 to 105	105 to 120	120 to 135	135 to 150	150 to 165	165 to 180	180 to 195	195 to 210	210 to 225	225 to 240	240 to 255	T'tls
15 to 17			1															1
13 to 15	1	1																2
11 to 13	5	1	1	2														9
9 to 11	13	8	0	4		2	2											29
7 to 9	11	15	10	5	7	2	6	7			1		1					65
5 to 7	4	13	16	22	10	7	11	7	1	2								93
3 to 5	12	13	12	12	25	12	5	10	4	2	2	3	1	1			1	115
1 to 3	4	8	18	18	20	16	32	14	8	5	3	3	2	1		1		153
− 1 to 1	5	4	7	34	21	24	27	29	8	10	7	5	1		2			184
− 3 to − 1	5	12	14	14	16	18	21	13	7	7	3	2		1	1	1		135
− 5 to − 3	2	17	8	18	12	18	17	10	6	1	5	1	1	1	1			118
− 7 to − 5	6	6	12	3	14	9	3	6	4	1			1					65
− 9 to − 7	11	6	9	3	9	2	1	2	2									45
−11 to − 9	3	6	3				1											13
−13 to −11	4	1	1		1													7
−15 to −13	1	1	1		1													4
−17 to −13																		
Totals	87	112	113	135	136	110	126	98	40	28	21	14	7	4	4	2	1	1038

Whence

$$(16)\qquad \beta_1 = \frac{(1+k)[-3m_1 - m_1^3 + k(3m_2\sigma^2 + m_2^3)]^2}{[1+m_1^2+k(\sigma^2+m_2^2)]^3},$$

$$(17)\qquad \beta_2 = \frac{(1+k)[3+6m_1^2+m_1^4+k(3\sigma^4+6m_2^2\sigma^2+m_2^4)]}{[1+m_1^2+k(\sigma^2+m_2^2)]^2}.$$

Thus, for any special population of the form (15), R' and ρ' can be easily calculated.

Samples of four were drawn from a population approximately represented by

$$(18)\qquad f_1(x) = 648\left[\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x+1.7)^2} + \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-1.7)^2}\right]$$

The actual sampled population is shown in Chart A and is hereinafter called Population I.

Table I shows the distribution of 1038 samples of four drawn from Population I with respect to the observed values of the means and the variances. The arrays for constant values of the variances are at first distinctly bimodal, gradually becoming unimodal. Chart I shows the means of arrays of Table I with the regression lines as calculated without correction for groupings. It is apparent that the locus of the mean variances for a given value of the means diverges a great deal from a straight line. This regression relation looks as though it was a normal curve,

which is what would be expected from (8) with $\sigma^2 - 1 = 0$. The theoretical and actual correlation coefficients for this and three subsequent tables are compared in Table V and the constants of the marginal distributions of Tables I to IV are presented in Table VI.

If the deviations of the means of the samples of Table I from the mean of Population I are squared, Table II results. Chart II shows the means of arrays and regression lines of Table II. The regression lines are very poor fits to the means of the arrays which are, apparently, exponential loci.

Table III shows the distribution of 1058 samples of four drawn from a population approximately represented by

$$(19)\quad f_2(x) = .972\left[\frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}(x+.8)^2} + \frac{1}{3\sqrt{2\pi}}\, e^{-\frac{1}{2}(x-2.4)^2}\right]$$

with respect to the observed values of the means and variances of the samples. The actual sampled population is presented in Chart B and is hereinafter called Population II. Chart III shows the means of arrays and regression lines of Table III. This chart resembles Chart I in that the locus of the mean variances for given values of the means is so obviously non-linear. Also, a glance at Table III is sufficient to see that the arrays vary markedly in skewness.

Table IV shows the relation between the means squared and variances of samples of four from Population II. Chart IV shows the means of arrays and regression lines for Table IV. In this case the regression relations seem to be fairly near linear, and the frequency distributions of the arrays do not change strikingly.

CHART I

The Means of Arrays and Regression Lines of the Means and Variances of Samples from Population I

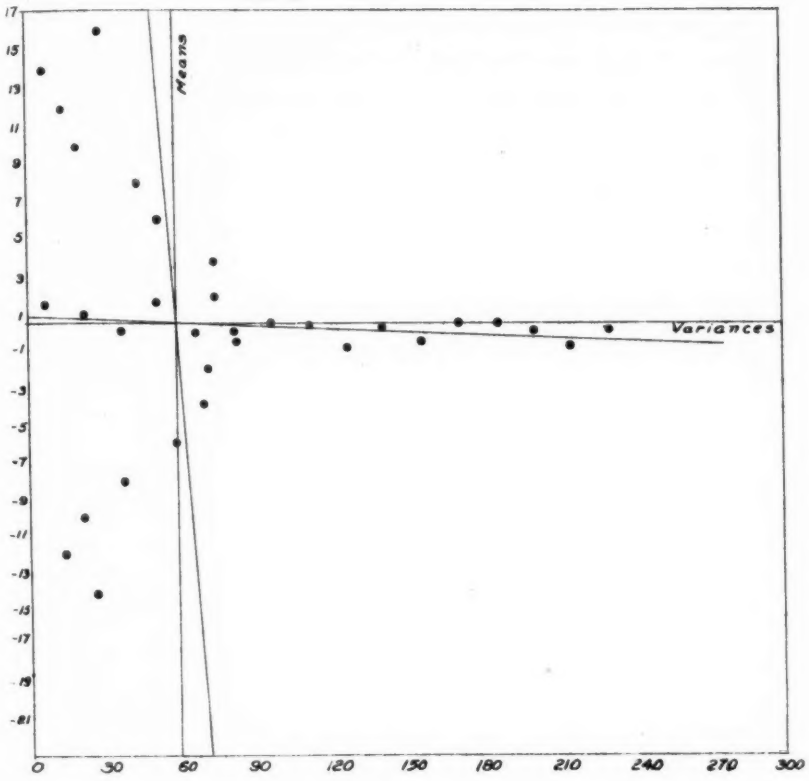


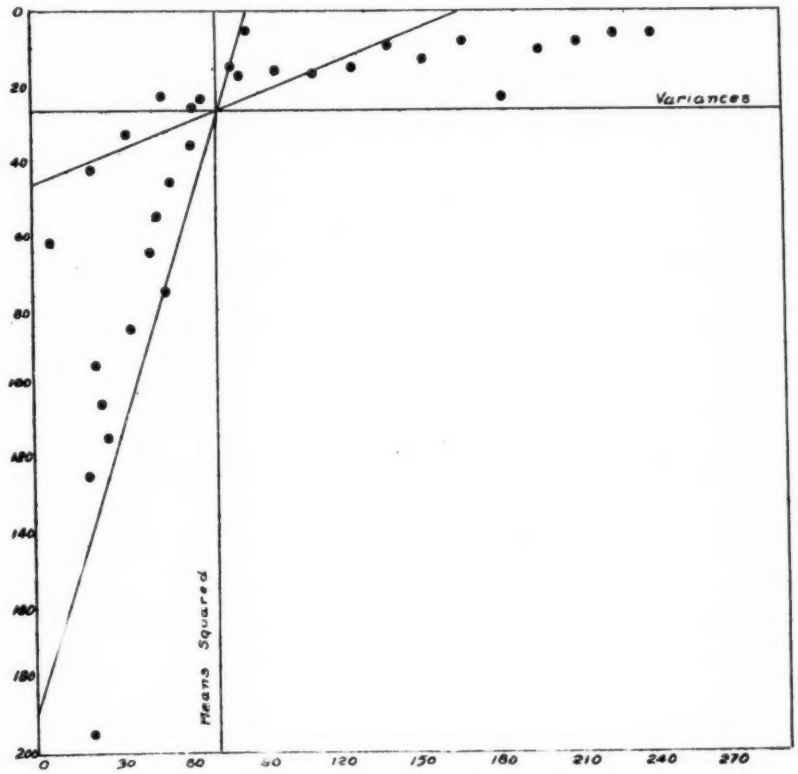
TABLE II

Correlation Table Showing the Relation Between the Means Squared and Variances of Samples of Four from Population I

Means Squared of Samples	*Variances of Samples*																	
	0 to 15	15 to 30	30 to 45	45 to 60	60 to 75	75 to 90	90 to 105	105 to 120	120 to 135	135 to 150	150 to 165	165 to 180	180 to 195	195 to 210	210 to 225	225 to 240	240 to 255	T'ls
0 to 10	13	25	42	64	59	58	83	56	25	22	12	10	3	2	3	2	1	480
10 to 20	7	21	10	18	24	19	14	15	6	3	8	4	2	2				154
20 to 30	5	14	13	21	16	15	11	9	2									106
30 to 40	4	9	13	12	17	9	6	5	2	3								80
40 to 50	7	7	11	6	5	4	2	5	3				1					51
50 to 60	7	5	6	2	5			4	2									31
60 to 70	10	6	6	4	6	2	5	2										41
70 to 80	6	6	3	2	1	1	2	2			1		1					25
80 to 90	5	3	1		2	2	1											14
90 to 100	5	3	2	2														12
100 to 110	4	6	1	2			1											14
110 to 120	3	1	1				1											6
120 to 130			1															2
130 to 140	5	2	2	1	1													11
140 to 150	2	1																3
150 to 160																		
160 to 170	1																	1
170 to 180		2	1															3
180 to 190	2																	2
190 to 200																		
220 to 230				1														1
250 to 260		1																1
Totals	87	112	113	135	136	110	126	98	40	28	21	14	7	4	4	2	1	1038

CHART II

The Means of Arrays and Regression Lines of the Means Squared and Variances of Samples from Population I



NOTE: The last thirteen class intervals of the means squared are grouped into one group.

CHART B

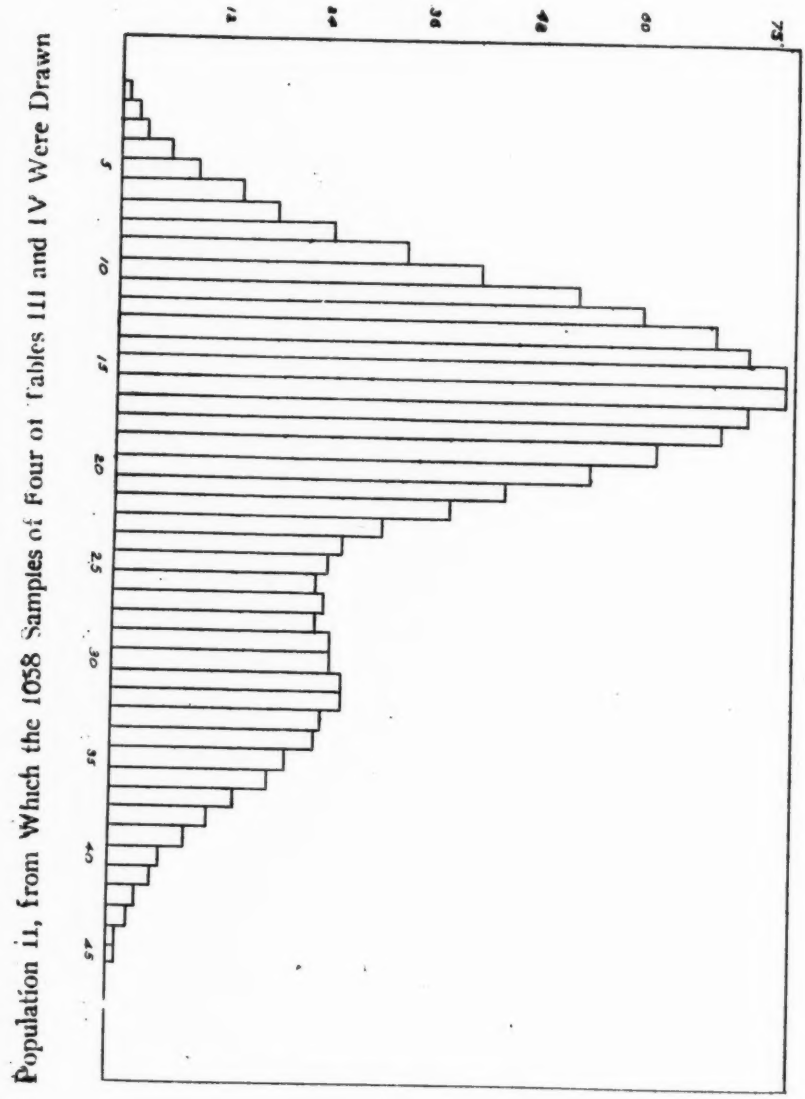

Population II, from Which the 1058 Samples of Four of Tables III and IV Were Drawn

TABLE III

Correlation Table Showing the Relation Between the Means and Variances of Samples of Four from Population II

Means of Samples	*Variances of Samples*																	
	0 to 15	15 to 30	30 to 45	45 to 60	60 to 75	75 to 90	90 to 105	105 to 120	120 to 135	135 to 150	150 to 165	165 to 180	180 to 195	195 to 210	210 to 225	225 to 240	240 to 255	T'ls
15 to 17	1		1															2
13 to 15		2	1															3
11 to 13			2				1	1										4
9 to 11	2	2	3		5	1	0	4					1					18
7 to 9	3	3	10	4	6	7	2	4	5	4		1						49
5 to 7	5	3	13	17	6	10	4	6	10	2	2	4			1			83
3 to 5	4	11	12	18	24	12	18	9	2	6	3	2						121
1 to 3	9	13	21	22	19	27	7	7	5	7	6		1		1	1		146
− 1 to 1	24	26	29	24	16	13	15	8	8	1	4	3	1		1			173
− 3 to − 1	45	36	26	22	14	14	11	4	5	1	1	1	1			1		182
− 5 to − 3	59	36	21	15	10	8			1	2								152
− 7 to − 5	36	29	7	7	2	1	1	1			1							85
− 9 to − 7	14	9	10		1													34
−11 to − 9	3	1																4
−13 to −11	1	1																2
−15 to −13																		
−17 to −15																		
Totals	206	172	156	129	103	93	59	44	36	23	17	11	4		3	2		1058

CHART III

The Means of Arrays and Regression Lines of the Means and Variances of Samples from Population II

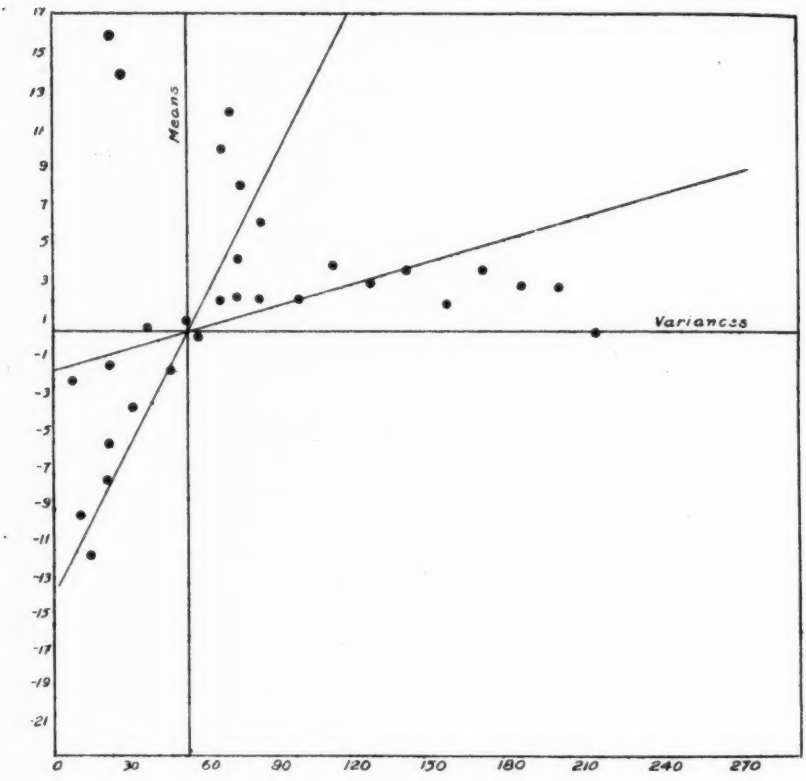


TABLE IV

Correlation Table Showing the Relation Between the Means Squared and Variances of Samples of Four from Population II

Means Squared of Samples \ *Variances of Samples*	0 to 15	15 to 30	30 to 45	45 to 60	60 to 75	75 to 90	90 to 105	105 to 120	120 to 135	135 to 150	150 to 165	165 to 180	180 to 195	195 to 210	210 to 225	225 to 240	240 to 255	T'ls
0 to 15	117	101	93	89	68	67	46	24	19	13	12	5	4		2	2		661
15 to 30	36	35	24	21	17	12	5	7	5	5	3	2						172
30 to 45	26	16	7	15	5	6	5	3	7		2	3			1			96
45 to 60	10	9	14	2	6	2	1	2	1	2	0	1						50
60 to 75	5	4	9	2	1	3	1	3	3	2								33
75 to 90	6	3	3		2	2		3	1	1								21
90 to 105	3		2		2	1		1					1					10
105 to 120	1	1			2													4
120 to 135	1																	2
135 to 150		1					1											2
150 to 165			2															2
165 to 180			1															1
180 to 195		1																1
195 to 210																		
210 to 225		1																1
225 to 240	1		1															2
240 to 255																		
Totals	206	172	156	129	103	93	59	44	36	23	17	11	4		3	2		1058

CHART IV

The Means of Arrays and Regression Lines of the Means Squared and Variances of Samples from Population II

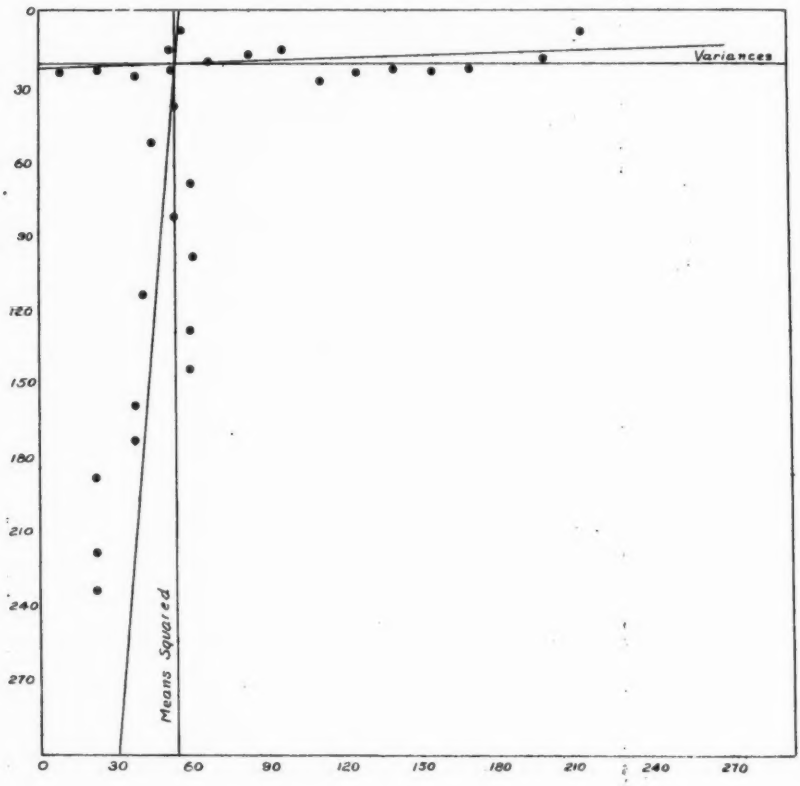


TABLE V

Correlation Coefficients of Tables I-IV

	Correlation-Coefficient	
Number of Table	Theoretical	Actual[1]
I	.00	−.05
II	−.34	−.37
III	.40	.37
IV	−.07	−.05

TABLE VI

Constants of the Marginal Distributions of Tables I-IV in Terms of Class Intervals

Marginal Distribution	Mean	Standard Deviation
Means of Samples from Population I	.252[2]	2.467
Variances of Samples from Population I	4.890[3]	2.900
Means Squared of Samples from Population I	3.591[3]	3.203
Means of Samples from Population II	.07	2.237
Variances of Samples from Population II	3.570[3]	2.854
Means Squared of Samples from Population II	1.408[3]	1.744

[1]Calculated without corrections for grouping.

[2]Is so far from zero because of the groupings employed. Many means were exactly odd integers. These were all put forward into higher classes, making the calculated mean too large.

[3]Origin taken at the beginning of the range.

From the results for the case of samples of two and from the results of empirical sampling, it seems clear that the simplest regression relation that is generally applicable to the means and variances, means squared and variances, of samples from populations which are the combinations of normal populations is parabolic. For small samples and for certain values of the parameters of the sampled population the regression relations may involve exponential terms that are quite important. As the size of the samples increases, it is expected that this exponential term will decrease in influence. It seems plausible that even with large samples the regression relation of means and variances, means squared and variances will remain essentially parabolic. It is not expected that the determination of a good approximation to the regression relations will serve to give an adequate notion of the probability relations of the means and variances, means squared and variances of samples from a population represented by (1), because the arrays may vary in number of modes, in skewness, in dispersion, and in other characteristics. For instance, surface (7) may be trimodal so that arrays may be bimodal or unimodal, and in such a case the arrays must vary markedly. Surfaces (7) and (10) with 2 replaced by n and with the terms suitably weighted are valuable approximations to the probability relations of the means and variances, means squared and variances of samples drawn from a population represented by (1).

G. A. Baker

A TABLE TO FACILITATE THE FITTING OF CERTAIN LOGISTIC CURVES

By

JOSHUA L. BAILEY, JR.

The most useful generalization of the logistic curve is that having the form

$$(1) \qquad y = \frac{k}{1+e^{a+bx+cx^2+gx^3}} \ldots .$$

In practice it will seldom be found necessary to use higher powers of x . This equation may also be written

$$(2) \qquad Y = a + bx + cx^2 + gx^3$$

in which $Y \equiv \log \frac{k-y}{y}$.

If we can evaluate the constant k with reasonable accuracy, the value of Y corresponding to each observed value of y can be computed, and then the values of the coefficients $a, b, c,$ and g , in equation (1) may be obtained by fitting equation (2) as a generalized parabola by the method of least squares.

The normal equations necessary to make this fit will be found to be

$$a\Sigma x^0 + b\Sigma x + c\Sigma x^2 + g\Sigma x^3 = \Sigma Y$$

$$a\Sigma x + b\Sigma x^2 + c\Sigma x^3 + g\Sigma x^4 = \Sigma x Y$$

$$a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4 + g\Sigma x^5 = \Sigma x^2 Y$$

$$a\Sigma x^3 + b\Sigma x^4 + c\Sigma x^5 + g\Sigma x^6 = \Sigma x^3 Y.$$

In the special case where the observations have been made at regular intervals (that is, where the successive values of x are in arithmetic progression) the solution of these normal equations may be greatly simplified. We may then select an arbitrary origin in the middle of the range of observations, so that for every positive value of x there will be a corresponding negative value of equal absolute magnitude. Thus the sums of the odd powers of x will all be zero.

If the number of observations be odd, the middle one will, of course, be chosen for the origin, and the unit of the scale will be the interval between successive values of x. If the number of observations be even, the origin will be midway between the middle pair of observations, and it will be found more convenient to take half the interval as scale unit. In the former case, x will take all integral values between $+n$ and $-n$, while in the latter case x may take only the odd integral values.

If we set the sums of the odd powers of x in the normal equations equal to zero, and solve them simultaneously, we derive the following formulae for the literal coefficients:

$$A = \frac{\Sigma Y \cdot \Sigma X^4 - \Sigma X^2 Y \cdot \Sigma X^2}{\Sigma X^4 \cdot \Sigma X^0 - (\Sigma X^2)^2}, \quad C = \frac{\Sigma X^2 Y \cdot \Sigma X^0 - \Sigma Y \cdot \Sigma X^2}{\Sigma X^4 \cdot \Sigma X^0 - (\Sigma X^2)^2},$$

$$B = \frac{\Sigma XY \cdot \Sigma X^6 - \Sigma X^3 Y \cdot \Sigma X^4}{\Sigma X^6 \cdot \Sigma X^2 - (\Sigma X^4)^2}, \quad G = \frac{\Sigma X^3 Y \cdot \Sigma X^2 - \Sigma XY \cdot \Sigma X^4}{\Sigma X^6 \cdot \Sigma X^2 - (\Sigma X^4)^2}.$$

The use of capital letters indicates that the equation has been referred to the arbitrary origin.

In these formulae the factors involving Y must be computed from the observations, but those in which X alone occurs may be tabulated for all convenient values of n. Since Y does not occur in the denominators at all, these may be tabulated in the same way.

TABLE TO BE USED WHEN THE NUMBER OF OBSERVATIONS IS ODD

n	Σx^0	Σx^2	Σx^4	Σx^6	$\Sigma x^4 \cdot \Sigma x^0 - (\Sigma x^2)^2$	$\Sigma x^6 \cdot \Sigma x^2 - (\Sigma x^4)^2$	$\Sigma x^4 \div \Sigma x^2$
1	3	2	2	2	2	0	1.0
2	5	10	34	130	70	144	3.4
3	7	28	196	1,588	588	6,048	6.0
4	9	60	708	9,780	2,772	85,536	11.8
5	11	110	1,958	41,030	9,438	679,536	17.8
6	13	182	4,550	134,342	26,026	3,747,744	25.0
7	15	280	9,352	369,640	61,880	16,039,296	33.4
8	17	408	17,544	893,928	131,784	56,930,688	43.0
9	19	570	30,666	1,956,810	257,754	174,978,144	53.8
10	21	770	50,666	3,956,810	471,086	479,700,144	65.8
11	23	1,012	79,948	7,499,932	814,660	1,198,248,480	79.0
12	25	1,300	121,420	13,471,900	1,345,500	2,770,653,600	93.4
13	27	1,638	178,542	23,125,518	2,137,590	6,002,352,720	109.0
14	29	2,030	255,374	38,184,590	3,284,946	12,298,837,824	125.8
15	31	2,480	356,624	60,965,840	4,904,944	24,014,605,824	143.8
16	33	2,992	487,696	94,520,272	7,141,904	44,957,265,408	163.0
17	35	3,570	654,738	142,795,410	10,170,930	81,097,765,056	183.4
18	37	4,218	864,690	210,819,858	14,202,006	141,549,364,944	205.0
19	39	4,940	1,125,332	304,911,620	19,484,348	239,891,292,576	227.8
20	41	5,740	1,445,332	432,911,620	26,311,012	395,928,108,576	251.8
21	43	6,622	1,834,294	604,443,862	35,023,758	637,992,775,728	277.0
22	45	7,590	2,302,806	831,203,670	46,018,170	1,005,920,381,664	303.4
23	47	8,648	2,862,488	1,127,275,448	59,749,032	1,554,840,524,160	331.0
24	49	9,800	3,526,040	1,509,481,400	76,735,960	2,359,959,638,400	359.8
25	51	11,050	4,307,290	1,997,762,650	97,569,290	3,522,530,138,400	389.8

TABLE TO BE USED WHEN THE NUMBER OF OBSERVATIONS IS EVEN

n	ΣX^0	ΣX^2	ΣX^4	ΣX^6	$\Sigma X^4 \cdot \Sigma X^0 - (\Sigma X^2)^2$	$\Sigma X^6 \cdot \Sigma X^2 - (\Sigma X^4)^2$	$\Sigma X^4 \div \Sigma X^2$
1	2	2	2	2	0	0	1.0
3	4	20	164	1,460	256	2,304	8.2
5	6	70	1,414	32,710	3,584	290,304	20.2
7	8	168	6,216	268,008	21,504	6,386,688	37.0
9	10	330	19,338	1,330,890	84,480	65,235,456	58.6
11	12	572	48,620	4,874,012	256,256	424,030,464	85.0
13	14	910	105,742	14,527,630	652,288	2,038,772,736	116.2
15	16	1,360	206,992	37,308,880	1,462,272	7,894,388,736	152.2
17	18	1,938	374,034	85,584,018	2,976,768	25,960,393,728	193.0
19	20	2,660	634,676	179,675,780	5,617,920	75,123,949,824	238.6
21	22	3,542	1,023,638	351,208,022	9,974,272	196,144,058,880	289.0
23	24	4,600	1,583,320	647,279,800	16,839,680	470,584,857,600	344.2
25	26	5,850	2,364,570	1,135,561,050	27,256,320	1,051,840,857,600	404.2
27	28	7,308	3,427,452	1,910,402,028	42,561,792	2,213,790,808,320	469.0
29	30	8,990	4,842,014	3,100,048,670	64,440,320	4,424,337,967,104	538.6
31	32	10,912	6,689,056	4,875,056,032	94,978,048	8,453,141,250,048	613.0
33	34	13,090	9,060,898	7,457,991,970	136,722,432	15,525,242,320,896	692.2
35	36	15,540	12,062,148	11,134,523,220	192,745,728	27,535,076,464,896	776.2
37	38	18,278	15,810,470	16,265,976,038	266,712,576	47,338,548,401,664	865.0
39	40	21,320	20,437,352	23,303,463,560	362,951,680	79,144,486,327,296	958.6
41	42	24,682	26,088,874	32,803,672,042	486,531,584	129,030,886,752,768	1,057.0
43	44	28,380	32,926,476	45,446,398,140	643,340,544	205,615,957,434,624	1,160.2
45	46	32,430	41,127,726	62,053,929,390	840,170,496	320,919,084,186,624	1,268.2
47	48	36,848	50,887,088	83,612,360,048	1,084,805,120	491,452,517,928,960	1,381.0
49	50	41,650	62,416,690	111,294,934,450	1,386,112,000	739,590,829,286,400	1,498.6

Finally, the sign of G is determined by the direction in which the curve approaches the asymptote $y = 0$, and this may readily be told by inspection. But it not infrequently happens that a slight error in one of the observations may be sufficient to give G the wrong sign. In this case the limits between which the observations were taken must be changed, or a new value of k must be tried, or the faulty observation must be adjusted by a smoothing formula. It is obviously important therefore that some means be provided for determining the sign of G before the values of the coefficients are determined.

The condition that G shall be negative is $\frac{\Sigma X^3 Y}{\Sigma X Y} > \frac{\Sigma X^4}{\Sigma X^2}$. The second term in this inequality may be tabulated in the same way. The accompanying tables show the values of the functions

$$\Sigma X^0,\ \Sigma X^2,\ \Sigma X^4,\ \Sigma X^6,\ \Sigma X^4 \cdot \Sigma X^0 - (\Sigma X^2)^2,$$
$$\Sigma X^6 \cdot \Sigma X^2 - (\Sigma X^4)^2 \text{ and } \Sigma X^4 \div \Sigma X^2$$

for all values of n from 0 to 25 when the number of observations is odd and from 0 to 49 when they are even.

In the preparation of these tables, my thanks are due to the Zoological Society of San Diego for the use of the facilities afforded by its research department.

Joshua L. Bailey Jr.

THE GENERALIZATION OF STUDENT'S RATIO*

By

HAROLD HOTELLING

The accuracy of an estimate of a normally distributed quantity is judged by reference to its variance, or rather, to an estimate of the variance based on the available sample. In 1908 "Student" examined the ratio of the mean to the standard deviation of a sample.[1] The distribution at which he arrived was obtained in a more rigorous manner in 1925 by R. A. Fisher,[2] who at the same time showed how to extend the application of the distribution beyond the problem of the significance of means, which had been its original object, and applied it to examine regression coefficients and other quantities obtained by least squares, testing not only the deviation of a statistic from a hypothetical value but also the difference between two statistics.

Let ξ be any linear function of normally and independently distributed observations of equal variance, and let s be the estimate of the standard error of ξ derived by the method of maximum likelihood. If we let t be the ratio to s of the deviation of ξ from its mathematical expectation, Fisher's result is that the probability that t lies between t_1 and t_2 is

---

*Presented at the meeting of the American Mathematical Society at Berkeley, April 11, 1931.

[1]Biometrika, vol. 6 (1908), p. 1.

[2]Applications of Student's Distribution, Metron, vol. 5 (1925), p. 90.

$$(1) \qquad \frac{1}{\sqrt{\pi n}} \; \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \int_{t_1}^{t_2} \frac{dt}{\left(1+t^2/n\right)^{\frac{n+1}{2}}}$$

where n is the number of degrees of freedom involved in the estimate s.

It is easy to see how this result may be extended to cases in which the variances of the observations are not equal but have known ratios and in which, instead of independence among the observations, we have a known system of intercorrelations. Indeed, we have only to replace the observations by a set of linear functions of them which are independently distributed with equal variance. By way of further extension beyond the cases discussed by Fisher, it may be remarked that the estimate of variance s^2 may be based on a body of data not involved in the calculation of ξ. Thus the accuracy of a physical measurement may be estimated by means of the dispersion among similar measurements on a different quantity.

A generalization of quite a different order is needed to test the simultaneous deviations of several quantities. This problem was raised by Karl Pearson in connection with the determination whether two groups of individuals do or do not belong to the same race, measurements of a number of organs or characters having been obtained for all the individuals. Several "coefficients of racial likeness" have been suggested by Pearson and by V. Romanovsky with a view to such biological uses. Romanovsky has made a careful study[1] of the sampling distributions, assuming in each case that the variates are independently and normally

[1] V. Romanovsky, On the criteria that two given samples belong to the same normal population (on the different coefficients of racial likeness), Metron, vol. 7 (1928), no. 3, pp. 3-46; K. Pearson, On the coefficient of racial likeness, Biometrika, vol. 18 (1926), pp. 105-118.

distributed. One of Romanovsky's most important results is the exact sampling distribution of L, a constant multiple of the sum of the squares of the values of t for the different variates. This distribution function is given by a somewhat complex infinite series. For large samples and numerous variates it slowly approximates to the normal form; for 500 individuals, Romanovsky considers that an adequate approach to normality requires that no fewer than 62 characters be measured in each individual. When it is remembered that all these characters must be entirely independent, and that it is usually hard to find as many as three independent characters, the difficulties in application will be apparent. To avoid these troubles, Romanovsky proposes a new coefficient of racial likeness, H, the average of the ratios of variances in the two samples for the several characters. He obtains the exact distribution of H, again as an infinite series, though it approaches normality more rapidly than the distribution of L. But H does not satisfy the need for a comparison between magnitudes of characters, since it concerns only their variabilities.

Joint comparisons of correlated variates, and variates of unknown correlations and standard deviations, are required not only for biologic purposes, but in a great variety of subjects. The eclipse and comparison star plates used in testing the Einstein deflection of light show deviations in right ascension and in declination; an exact calculation of probability combining the two least-square solutions is desirable. The comparison of the prices of a list of commodities at two times, with a view to discovering whether the changes are more than can reasonably be ascribed to ordinary fluctuation, is a problem dealt with only very crudely by means of index numbers, and is one of many examples of the need for such a coefficient as is now proposed. We shall generalize Student's distribution to take account of such cases.

We consider p variates $x_1, x_2, \ldots, x_p$, each of which is measured for N individuals, and denote by $X_{i\alpha}$ the value of x_i for the α th individual. Taking first the problem

of the significance of the deviations from a hypothetical set of mean values $m_1, m_2, \ldots, m_p$, we calculate the means $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_p$, of the samples, and put

$$\xi_i = (\bar{x}_i - m_i)\sqrt{N}.$$

Then the mean values of the ξ_i will all be zero, and the variances and covariances will be the same as for the corresponding x_i, since the individuals are supposed chosen independently from an infinite population.[1] In order to estimate them with the help of the deviations

$$x_i = X_{i\alpha} - \bar{x}_i$$

from the respective means, we call $n = N - 1$ the number of degrees of freedom and take as the estimates of the variances and covariances,

$$a_{ji} = a_{ij} = \frac{1}{n}\sum_{\alpha=1}^{N} x_{i\alpha}\, x_{j\alpha} \qquad (2)$$

We next put:

$$\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{vmatrix}$$

[1] "Mean values" used in the sense of mathematical expectation: the variable quantity whose mean value is zero is defined as the expectation of its square; the covariance of two such quantities is the expectation of their product. Thus the correlation of the two in a hypothetical infinite population is the ratio of their covariance to the geometric mean of the variances.

$$A_{ij} = A_{ji} = \frac{\text{cofactor of } a_{ij} \text{ in } a}{a} \tag{3}$$

The measure of simultaneous deviations which we shall employ is

$$T^2 = \sum_{i=1}^{p} \sum_{j=1}^{p} A_{ij} \xi_i \xi_j . \tag{4}$$

For a single variate it is natural to take $A_{11} = 1/a_{11}$; then T reduces to t , the ordinary "critical ratio" of a deviation in a mean to its estimated standard error, a ratio which has "Student's distribution," (1). For examining the deviations from zero of two variates x and y ,

$$T = \frac{N}{1-r^2} \left\{ \frac{\bar{x}^2}{s_1^2} - \frac{2r\bar{x}\bar{y}}{s_1 s_2} + \frac{\bar{y}^2}{s_2^2} \right\},$$

where

$$s_1^2 = \frac{\Sigma (X-\bar{x})^2}{N-1}, \qquad s_2^2 = \frac{\Sigma (Y-\bar{y})^2}{N-1},$$

$$r = \frac{\Sigma (X-\bar{x})(Y-\bar{y})}{\sqrt{\Sigma (X-\bar{x})^2 \Sigma (Y-\bar{y})^2}}$$

For comparing the means of two samples, one of N_1 and the other of N_2 individuals, we distinguish symbols pertaining to the second sample by primes, and write

$$(5) \qquad \xi_i = \frac{\bar{x}_i - \bar{x}_i'}{\sqrt{1/N_1 + 1/N_2}}$$

$$n = N_1 + N_2 - 2,$$

$$(6) \qquad a_{ij} = \frac{1}{n}\left[\Sigma(X_{i\alpha} - \bar{x}_i)(X_{j\alpha} - \bar{x}_j) + \Sigma(X'_{i\alpha} - \bar{x}'_i)(X'_{j\alpha} - \bar{x}'_j)\right]$$

$$= \frac{1}{n}\left[\Sigma X_{i\alpha} X_{j\alpha} - N_1 \bar{x}_i \bar{x}_j + \Sigma X_{i\alpha} X_{j\alpha} - N_2 \bar{x}'_i \bar{x}'_j\right]$$

and take as our "coefficients of racial likeness" the value (4) of T^2, in which the ξ_i are calculated from (5) and the A_{ij} from (6) and (3).

Other situations to which the measure T^2 of simultaneous deviations can be applied include comparisons of regression coefficients and slopes of lines of secular trend, comparisons which for single variates have been explained by R. A. Fisher.[1] In each case we deal for each variate with a linear function ξ_i of the observed values, such that the sum of the squares of the coefficients is unity, so that the variance is the same as for a single observation, and such that the expectation of ξ_i is, on the hypothesis to be tested, zero. Deviations $x_{i\alpha}$ of the observations from means, or from trend lines or other such estimates, are used to provide the estimated variances and covariances a_{ij} by (2). The number of degrees of freedom n is the difference between the number N of individuals and the number q of independent linear relations which must be satisfied by the quan-

[1]Metron, loc. cit., and Statistical Methods for Research Workers, Oliver and Boyd, third edition (1928).

tities $x_{i1}, x_{i2}, \ldots, x_{iN}$ on account of their method of derivation. For all the variates, these relations and n must be the same.

The general procedure is to set up what may be called normal values $\bar{x}_{i\alpha}$ for the respective $X_{i\alpha}$, putting

$$x_{i\alpha} = X_{i\alpha} - \bar{x}_{i\alpha}. \tag{7}$$

The underlying assumption is that $X_{i\alpha}$ is composed of two parts, of which one, $\varepsilon_{i\alpha}$, is normally and independently distributed about zero with variance σ_i^2 which is the same for all the observations on x_i. The other component is determined by the time, place, or other circumstances of the α'th observation in some regular manner, the same for all the variates. Denoting this part by $\eta_{i\alpha}$, we have

$$X_{i\alpha} = \eta_{i\alpha} + \varepsilon_{i\alpha}.$$

Specifically, we take $\eta_{i\alpha}$ to be a linear function, with known coefficients $g_{\alpha s}$, of q unknown parameters $\zeta_{i1}, \ldots, \zeta_{iq}$ where $q < N$:

$$\eta_{i\alpha} = \sum_{s=1}^{q} g_{\alpha s}\, \zeta_{is}. \tag{8}$$

Thus in dealing with a secular trend representable by a polynomial in the time, we may take the g's as powers of the time-variable, the ζ's as the coefficients. For differences of means, the g's are 0's and 1's, and the ζ's the true means.

We estimate the ζ's by minimizing

$$2V_i = \sum_{\alpha=1}^{N} \varepsilon_{i\alpha}^2 = \sum_{\alpha=1}^{N} (X_{i\alpha} - \eta_{i\alpha})^2. \tag{9}$$

Substituting from (8), differentiating with respect to ζ_{is}, and replacing $\eta_{i\alpha}$ by $\bar{x}_{i\alpha}$ for the minimizing value, we obtain:

$$(10) \quad \sum_{\alpha=1}^{N} g_{\alpha s}(X_{i\alpha} - \bar{x}_{i\alpha}) = 0, \qquad (s = 1, 2, \cdots, q)$$

or by (7),

$$(11) \quad \sum_{\alpha=1}^{N} g_{\alpha s}\, x_{i\alpha} = 0 \qquad (s = 1, 2, \cdots, q)$$

Denoting also the minimizing values of ζ_{is} by z_{is}, we have made from (8),

$$\bar{x}_{i\alpha} = \sum_{s=1}^{q} g_{\alpha s}\, z_{is}$$

Subtracting (8),

$$(12) \qquad \bar{x}_{i\alpha} - \eta_{i\alpha} = \sum_{s=1}^{q} g_{\alpha s}(z_{is} - \zeta_{is})$$

From (9),

$$2V = \sum_{\alpha=1}^{N} \left[(X_{i\alpha} - \bar{x}_{i\alpha}) + (\bar{x}_{i\alpha} - \eta_{i\alpha})\right]^2$$

$$(13) \qquad = \sum_{\alpha=1}^{N} (X_{i\alpha} - \bar{x}_{i\alpha})^2 + 2\sum_{\alpha=1}^{N} (X_{i\alpha} - \bar{x}_{i\alpha})(\bar{x}_{i\alpha} - \eta_{i\alpha})$$

$$+ \sum_{\alpha=1}^{N} (\bar{x}_{i\alpha} - \eta_{i\alpha})^2$$

The middle term, by (12), equals

$$2\sum_{\alpha=1}^{N}\sum_{s=1}^{q} g_{\alpha s}(X_{i\alpha} - \bar{x}_{i\alpha})(z_{is} - \zeta_{is}),$$

this, by (10), is zero. Hence, by (7) and (13),

$$U_i = V_i + W_i ,$$

where

$$2V_i = \sum_{\alpha=1}^{N} x_{i\alpha}^2$$

$$2W_i = \sum_{\alpha=1}^{N} (\bar{x}_{i\alpha} - \eta_{i\alpha})^2$$

If the q equations (10) be solved for $\bar{x}_{i1}$, $\bar{x}_{i2}$, . . . , $\bar{x}_{iN}$, the values of these quantities will be found to be homogeneous linear functions of the observations $X_{i\alpha}$. By (7), therefore, the quantities

$$\bar{x}_{i1} , \bar{x}_{i2} , \ldots\ldots , \bar{x}_{iN}$$

are homogeneous linear functions of the $X_{i\alpha}$. But they are not linearly independent functions, since they are connected by the q relations (11). Hence V is a quadratic form of rank

$$n = N - q ..$$

Since V_i , by (9), is of rank N, W is of rank q.

This shows that Np new quantities $x'_{i\alpha}$, given by equations of the form

$$x'_{i\alpha} = \sum_{\beta=1}^{N} c_{\alpha\beta} x_{i\beta} = \sum_{\beta=1}^{N} C_{\alpha\beta} X_{i\beta} , (\alpha = 1, 2, \cdots, n)$$

(14)

$$x'_{i\alpha} = \sum_{\beta=1}^{N} c_{\alpha\beta} (\bar{x}_{i\beta} - \eta_{i\beta}) = \sum_{\beta=1}^{N} (C_{\alpha\beta} X_{i\beta} - c_{\alpha\beta} \eta_{i\beta}) , (\alpha = n+1, \cdots, N)$$

can be found such that

$$2V_i = \sum_{\alpha=1}^{N} x_{i\alpha}^2 = \sum_{\alpha=1}^{N} x'^2_{i\alpha} ,$$

(15)

$$2W_i = \sum_{\alpha=n+1}^{N} x'^2_{i\alpha} ,$$

and therefore

$$2U_i = \sum_{\alpha=1}^{N} x'^2_{i\alpha} . \tag{16}$$

Substituting (14) in (15) and equating like coefficients,

$$\sum_{\alpha=1}^{n} c_{\alpha\beta}\, c_{\alpha\gamma} = \delta_{\beta\gamma} \tag{17}$$

where $\delta_{\beta\gamma}$ is the Kronecker delta, equal to 1 if $\beta = \gamma$, to 0 if $\beta \neq \gamma$

The coefficients $c_{\alpha\beta}$ depend only on the $g_{\alpha s}$, which have been assumed to be the same for all the p variates. Thus (14) may be written

$$x'_{j\alpha} = \sum_{\gamma=1}^{N} c_{\alpha\gamma}\, x_{j\gamma} .$$

Multiplying by (14), summing with respect to α from 1 to n, and using (17),

$$\sum_{\alpha=1}^{n} x'_{i\alpha}\, x'_{j\alpha} = \sum_{\alpha=1}^{n} \sum_{\beta=1}^{N} \sum_{\gamma=1}^{N} c_{\alpha\beta}\, c_{\alpha\gamma}\, x_{i\beta}\, x_{j\gamma}$$

(18)

$$= \sum_{\beta=1}^{N} \sum_{\gamma=1}^{N} \delta_{\beta\gamma}\, x_{i\beta}\, x_{j\gamma} = \sum_{\beta=1}^{N} x_{i\beta}\, x_{j\beta}$$

Just as in (2), we define a_{ij} in this generalized case by

$$a_{ij} = \frac{1}{n} \sum_{\alpha=1}^{N} x_{i\alpha} x_{j\alpha} . \tag{19}$$

Then by (18),

$$a_{ij} = \frac{1}{n} \sum_{\alpha=1}^{N} x'_{i\alpha} x'_{j\alpha} . \tag{20}$$

Of the last equation, (6) is a special case.

The random parts $\varepsilon_{i\alpha}$ of the observations on x_i have by hypothesis the distribution

$$\frac{1}{(\sigma_i \sqrt{2\pi})^N} e^{-U_i/2\sigma_i^2} d\varepsilon_{i1} d\varepsilon_{i2} \cdots d\varepsilon_{iN},$$

where V_i is given by (9). From what has been shown, it is clear that this may be transformed into

$$\frac{1}{(\sigma_i \sqrt{2\pi})^N} e^{-(x'^2_{i1} + x'^2_{i2} + \cdots + x'^2_{iN})/2\sigma_i^2} dx'_{i1} \cdots dx'_{iN},$$

showing that $x'_{i1}, \ldots, x'_{iN}$ are normally and independently distributed with equal variance σ_i^2.

The statistic ξ_i must be independent of the quantities $x'_{i1}, x'_{i2}, \ldots, x'_{in}$ entering into (20), its mean value must be zero, and its variance must be σ_i^2. These conditions are satisfied in the cases which have been mentioned, and are satisfied in general if ξ_i is a linear homogeneous function of $x'_{i,n+1}, \ldots, x'_{iN}$ with the sum of the squares of the coefficients equal to unity.

The measure of simultaneous discrepancy is

$$T^2 = \sum_{i=1}^{p} \sum_{j=1}^{p} A_{ij} \xi_i \xi_j ,$$

A_{ij} being defined by (3) on the basis of (19). It is evident that

$$(21)\quad T^2 = -\frac{\begin{vmatrix} 0 & \xi_1 & \xi_2 & \cdots & \xi_p \\ \xi_1 & a_{11} & a_{12} & \cdots & a_{1p} \\ \xi_2 & a_{21} & a_{22} & \cdots & a_{2p} \\ \cdots & a & \cdots & \cdots & \cdots \\ \xi_p & a_{p1} & a_{p2} & \cdots & a_{pp} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{vmatrix}}$$

as appears when the numerator is expanded by the first row, and the resulting determinants by their first columns.

A most important property of T is that it is an absolute invariant under all homogeneous linear transformations of the variates $x_i, \ldots, x_p$. This may be seen most simply by tensor analysis; for ξ_i is covariant of the first order and A_{ij} is contravariant of the second order.

The invariance of T shows that in seeking its sampling distribution we may, without loss of generality, assume that the variates $x_1, \ldots, x_p$ have, in the normal population, zero correlations and equal variances for they may always by a linear transformation be replaced by such variates.

Let us now take

$$\xi_i, x'_{i1}, x'_{i2}, \cdots, x'_{in}$$

as rectangular coordinates of a point P_i in space V_{n+1} of $n+1$ dimensions. Since these quantities are normally and independently distributed with equal variance about zero, the probability density for P_i has spherical symmetry about the origin. Indefinite repetition of the sampling would result in a globular cluster of representative points for each variate. Actually the sample in hand fixes the points $P_1, P_2, \ldots, P_p$, which may be regarded as taken independently.

We shall now show that T is a function of the angle θ between the ξ-axis and the flat space V_p containing the points $P_1, P_2, \ldots, P_p$ and the origin 0. We shall denote by A the point on the ξ-axis of coordinates 1, 0, 0, . . ., 0, and by V_n the flat space containing the remaining axes. Since in V_{n+1} one equation specifies V_n and $n+1-p$ equations V_p, the intersection of V_n and V_p is specified by all these $n+2-p$ equations, and is therefore of $p-1$ dimensions. Call it V_{p-1}.

If $P_1, P_2, \ldots, P_p$ be moved about in V_p, θ will not change, and neither will T, since T is invariant under linear transformations, equivalent to such motions of the P_i. Hence T always has the value which it takes if all the lines OP_1, OP_2, . . ., OP_p are perpendicular, with the last $p-1$ of these lines lying in V_{p-1}. In this case the angle AOP_1 equals θ. Applying to the coordinates of A and of P_1 the formula for the cosine of an angle at the origin of lines to $(x_1, x_2, \ldots)$ and $(y_1, y_2, \ldots)$, namely,

$$\cos\theta = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}} \tag{22}$$

We obtain

$$\cos\theta = \frac{\xi}{\sqrt{\xi_1^2 + x_{11}'^2 + \cdots + x_{1n}'^2}}.$$

Since $x_{11}'^2 + \cdots + x_{1n}'^2 = na_{11}$,

it follows that

$$(23) \qquad n \cot^2 \theta = \xi_1^2 / a_{11}.$$

The fact that $P_2, P_3, \ldots, P_p$ lie in V_{p-1}, and therefore in V_n, shows that in this case

$$\xi_2 = \xi_3 = \cdots = \xi_p = 0.$$

Because $OP_1, OP_2, \ldots, OP_p'$ are mutually perpendicular, (20) and (22) show that $a_{ij} = 0$ whenever $i \neq j$. Hence, by (21) and (23),

$$(24) \qquad T = \xi_1 / a_{11} = \sqrt{n} \cot \theta.$$

By this result the problem of the sampling distribution of T is reduced to that of the angle θ between a line OA in V_{n+1} and the flat space V_p containing p other lines drawn independently through the origin. The distribution will be unaffected if we suppose V_p fixed and OA drawn at random, with spherical symmetry for the points A.[1] Let us then, abandoning the coordinates hitherto used, take new axes of rectangular coordinates $y_1, y_2, \ldots, y_{n+1}$, of which the first p lie in V_p. A unit hypersphere about 0 is defined in terms of the general-

[1]This geometrical interpretation of T shows its affinity with the multiple correlation coefficient, whose interpretation as the cosine of an angle of a random line with a V_p enabled R. A. Fisher to obtain its exact distribution (Phil. Trans., vol. 213B, 1924, p. 91; and Proc. Roy. Soc., vol. 121A, 1928, p. 654). The omitted steps in Fisher's argument may be supplied with the help of generalized polar coordinates as in the text. Other examples of the use of these coordinates in statistics have been given by the author in The Distribution of Correlation Ratios Calculated from Random Data, Proc. Nat. Acad. Sci., vol. 11 (1925), p. 657, and in The Physical State of Protoplasm, Koninklijke Akademie van Wetenschappen te Amsterdam, verhandlingen, vol. 25 (1928), no. 5, pp. 28-31.

ized latitude-longitude parameters $\phi_1, \ldots, \phi_n$ if we put

$$
\begin{aligned}
y_1 &= \sin\phi_1 \sin\phi_2 \sin\phi_3 \cdots \sin\phi_{p-1} \cos\phi_p \\
y_2 &= \cos\phi_1 \sin\phi_2 \sin\phi_3 \cdots \sin\phi_{p-1} \cos\phi_p \\
y_3 &= \cos\phi_2 \sin\phi_3 \cdots \sin\phi_{p-1} \cos\phi_p \\
y_4 &= \cos\phi_3 \cdots \sin\phi_{p-1} \cos\phi_p \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
y_p &= \cos\phi_{p-1} \cos\phi_p \\
y_{p+1} &= \sin\phi_p \cos\phi_{p+1} \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
y_n &= \sin\phi_p \sin\phi_{p+1} \cdots \cos\phi_n \\
y_{n+1} &= \sin\phi_p \sin\phi_{p+1} \cdots \sin\phi_n,
\end{aligned}
$$

for the sum of the squares is unity. Since

$$y_{p+1}^2 + \cdots + y_{n+1}^2 = \sin^2\phi_p$$

we have

$$\phi_p = \theta \ .$$

The element of probability is proportional to the element of generalized area, which is given by

$$\sqrt{D}\, d\phi_1\, d\phi_2 \cdots d\phi_n,$$

where D is an n-rowed determinant in which the element in the ith row and jth column is

$$\sum_{k=1}^{n+1} \frac{\partial y_k}{\partial \phi_i} \cdot \frac{\partial y_k}{\partial \phi_j}$$

For $i \neq j$, this is zero. Of the diagonal elements, the first p-1 contain the factor $\cos^2 \phi_p$; the pth is unity; and the remaining $n-p$ elements contain the factor $\sin^2 \phi_p$. Since ϕ is not otherwise involved, the element of area is the product of

$$\cos^{p-1} \phi_p \sin^{n-p} \phi_p \, d\phi_p$$

by factors independent of ϕ_p. The distribution function of θ is obtained by replacing ϕ_p by θ and integrating with respect to the other parameters. Since θ lies between 0 and $\pi/2$, we divide by the integral between these limits and obtain for the frequency element,

$$\frac{2\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{n-p+1}{2}\right)} \cos^{p-1}\theta \sin^{n-p}\theta \, d\theta.$$

Substituting from (24) we have as the distribution of T:

$$(25) \qquad \frac{2\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{n-p+1}{2}\right) n^{p/2}} \frac{T^{p-1} dT}{\left(1+\frac{T^2}{n}\right)^{\frac{n+1}{2}}}$$

For $p = 1$ this reduces to the form of Student's distribution given by Fisher and tabulated in the issue of Metron cited; however, as T may be negative as well as positive in this case, Fisher omits the factor 2.

For $p = 2$ the distribution becomes

$$\frac{n-1}{n} \frac{T\,dT}{\left(1+\frac{T^2}{n}\right)^{\frac{n+1}{2}}}.$$

From this it is easy to calculate as the probability that a given value of T will be exceeded by chance,

$$(26) \qquad P = \frac{1}{\left(1+\frac{T^2}{n}\right)^{\frac{n-1}{2}}}$$

a very convenient expression.

The probability integral for higher values of p may be calculated in various ways, the most direct being successive integration by parts, giving a series of terms analogous to (26) to which, if p is odd, is added an integral which may be evaluated with the help of the tables of Student's distribution. If p is large, this process is laborious; but other methods are available.

The probability integral is reduced to the incomplete beta function if we put

$$x = (1 + T^2/n)^{-1},$$

for then the integral of (25) from T to infinity becomes

$$P = I_x\left(\frac{n-p+1}{2}, \frac{p}{2}\right),$$

the notation being

$$B_x(p,q) = \int_0^x x^{p-1}(1-x)^{q-1}dx,$$

$$B(p,q) = \int_0^1 x^{p-1}(1-x)^{q-1}dx,$$

$$I_x(p,q) = \frac{B_x(p,q)}{B(p,q)}.$$

Many methods of calculation have been discussed by H. E. Soper[1] and by V. Romanovsky.[2] An extensive table of the incomplete beta function being prepared under the supervision of Professor Karl Pearson has not yet been published.

Perhaps the most generally useful method now available is

[1]Tracts for Computers, no. 7 (1921).

[2]On certain expansions in series of polynomials of incomplete B-functions (in English), Recueil Math. de la Soc. de Moscou, vol. 33 (1926), pp. 207-229.

to make the substitution

$$z = \frac{1}{2} \log_e (n-p+1) T^2 - \frac{1}{2} \log_e np,$$

$$n_1 = p$$

$$n_2 = n-p+1,$$

reducing (25) to a form considered by Fisher. Table VI in his book, Statistical Methods for Research Workers, gives the values of z which will be exceeded by chance in 5 per cent and in 1 per cent of cases. If the value of z obtained from the data is greater than that in Fisher's table, the indication is that the deviations measured are real.

If the variances and covariances are known a priori, they are to be used instead of the a_{ij} ; the resulting expression $-T$ has the well known distribution of χ, with p degrees of freedom. For very large samples the estimates of the covariances from the sample are sufficiently accurate to permit the use of the χ distribution for T. This is well shown by (25), in which, as n increases, the factor involving T approaches

$$T^{p-1} e^{-T^2/2} dT,$$

which is proportional to the frequency element for χ when χ is put for T.

As Pearson pointed out, the labor of calculating χ, which we replace by T, is prohibitive when forty or fifty characters are measured on each individual. With two, three, or four characters, however, the labor is very moderate, and the results far more accurate than any attainable with the Pearson coefficient. The great advantage of using T is the simplicity of its distribution, with its complete independence of any correlations among the variates which may exist in the population.

To means of a single variate it is customary to attach a

"probable error," with the assumption that the difference between the true and calculated values is almost certainly less than a certain multiple of the probable error. A more precise way to follow out this assumption would be to adopt some definite level of probability, say $P = .05$, of a greater discrepancy, and to determine from a table of Student's distribution the corresponding value of t, which will depend on n; adding and subtracting the product of this value of t by the estimated standard error would give upper and lower limits between which the true values may with the given degree of confidence be said to lie. With T an exactly analogous procedure may be followed, resulting in the determination of an ellipse or ellipsoid centered at the point ξ_1, ξ_2, . . ., ξ_p. Confidence corresponding to the adopted probability P may then be placed in the proposition that the set of true values is represented by a point within this boundary.

Harold Hotelling

SYSTEMS OF POLYNOMIALS CONNECTED WITH THE CHARLIER EXPANSIONS AND THE PEARSON DIFFERENTIAL AND DIFFERENCE EQUATIONS*

By

EMANUEL HENRY HILDEBRANDT

INTRODUCTION

The problem of fitting mathematical curves to statistical data has commanded the attention of statisticians and mathematicians for many years. The curves referred to the most by English-speaking biometricians and mathematicians are perhaps those developed by Pearson from 1895-1916.[1] He showed that a series of curves could be obtained by assigning various values to the parameters in a certain first order differential equation. A few years later, Charlier[2], attacking the same question from a differ-

---

*A dissertation submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in the University of Michigan—August, 1931.

[1]Karl Pearson, "Mathematical Contributions to the Theory of Evolution," Philosophical Transactions, A, Vol. 186 (1895), pp. 343-414; also "Supplement to a Memoir on Skew Variation," Phil. Trans., Vol. 197 (1901), pp. 443-456; also "Second Supplement to a Memoir on Skew Variation," Phil. Trans., A, Vol. 216 (1916), pp. 429-457.

[2]C. V. L. Charlier, "Ueber das Fehlergesetz," Arkiv for Matematik, Astronomi och Fysik, Vol. 2, No. 8 (1905), pp. 1-9; also "Ueber die Darstellung willkuerlicher Funktionen," Arkiv for Matematik, Astronomi och Fysik, Vol. 2, No. 20 (1905), pp. 1-35.

ent angle, showed that any function could probably be approximated by using a certain function and its derivatives in the terms of the series:

$$F(x)=A_0 f(x)+A_1 f'(x)+A_2 f''(x)+\cdots$$

where the A_i are constants.

Charlier found that the constants A_n could be formally determined, the n th constant A_n being dependent on the moments of $F(x)$ of order not greater than n. He illustrated the method of procedure for the case where $y=f(x)$ was the equation of the normal curve of error, i. e. one of the Pearson curves. In fact, the successive derivatives of this particular function gave rise to a well known system of polynomials, namely the Hermite polynomials, and the coefficients are dependent upon these polynomials also.

In recent years, Romanovsky[1] has succeeded in obtaining similar results for the case in which some of the other of the Pearson curves are used as the $f(x)$ in the Gram-Charlier series. The successive derivatives of these other special Pearson type curve functions also result in systems of polynomials which bear fundamental relations to each other.

It is the object of this investigation to show:

(1) That the constants obtained by Charlier for his Type A series can be much more readily obtained by making use of certain existing biorthogonality conditions;

(2) That if the Type A series be generalized to the form:

$$F(x)=C_0 Q(x) f_0(x)+C_1 \frac{d}{dx} Q(x) f_1(x)+C_2 \frac{d^2}{dx^2} Q(x) f_2(x)+\cdots$$

---

[1] V. Romanovsky, "Generalization of some types of the frequency curves of Professor Pearson," Biometrika, Vol. 16 (1924), pp. 106-117; also "Sur quelques classes nouvelles de Polynomes orthogonaux," Comptes Rendus de L'Academie des Sciences, Vol. 188 (1929), pp. 1023-1025.

where $f_n(x)$ is a polynomial of degree n in x, then the C_n can also be formally determined and depend upon the moments of $F(x)$ of order at most n;

(3) That the form of the polynomials obtained by Charlier and Romanovsky for certain solutions of the Pearson differential equation can be found for any solution of this equation and that the relations existing between polynomials of the same system can also be generalized for the general solution and for the most part obtained without having the explicit form of the solution;

(4) That results analogous to those obtained in (1) and (3) can be derived for the Charlier Type B series and the analogue of Pearson's differential equation, finite differences replacing the derivative.

The writer wishes to particularly express his appreciation to Prof. H. C. Carver for the valuable aid he has given both in the stimulating instruction characterized by frankness in indicating unsolved problems in his classes and through direct suggestions in the preparation of this paper.

CHAPTER I

POLYNOMIALS CONNECTED WITH THE GRAM-CHARLIER SERIES

1. In the articles entitled "Ueber das Fehlergesetz" and "Ueber die Darstellung willkürlicher Funktionen"[1] Charlier proves the following well known theorem:

CHARLIER'S THEOREM FOR SERIES OF TYPE A—*If* $F(x)$ *is any real valued function of* x, *which has finite moments of all orders, then* $F(x)$ *may be formally expressed in terms of another function* $f(x)$ *and its derivatives as follows:*

$$(A)\quad F(x) = A_0 f(x) + A_1 f'(x) + A_2 f''(x) + \cdots + A_n f^{(n)}(x) + \cdots$$

where $f(x)$ *has the following properties:*

(a) $f(x)$ *and its derivatives are continuous for all real values of* x,

(b) $f(x)$ *and its derivatives vanish for* $x = +\infty$ *and* $-\infty$

(c) $\lim_{x=\pm\infty} x^m f^{(n)}(x) \quad 0$ *for all* m *and* n,

(d) $\int_{-\infty}^{+\infty} f(x)\,dx \neq 0.$

The conditions (c) and (d) are not given in Charlier's articles, but an eaxmination of the proof shows that he assumes implicitly that they are satisfied. $f(x) = \frac{x}{1+x^2}$ satisfies (a) and (b) without satisfying (c) and (d).

In the first section of the latter paper, Charlier determines the constants $A_0, A_1, A_2, \ldots, A_n, \ldots$. He takes the series (A), multiplies it successively by $1, x, x^2, \ldots$, and integrates each result between the limits $-\infty$ to $+\infty$. The fol-

[1] C. V. L. Charlier, loc. cit.

lowing equations result:

$$\int_{-\infty}^{+\infty} F(x)dx = A_0 \int_{-\infty}^{+\infty} f(x)dx$$

$$\int_{-\infty}^{+\infty} x\,F(x)dx = A_0 \int_{-\infty}^{+\infty} x\,f(x)dx + A_1 \int_{-\infty}^{+\infty} x\,f'(x)\,dx$$

$$\int_{-\infty}^{+\infty} x^2 F(x)dx = A_0 \int_{-\infty}^{+\infty} x^2 f(x)dx + A_1 \int_{-\infty}^{+\infty} x^2 f'(x)dx + A_2 \int_{-\infty}^{+\infty} x^2 f''(x)dx$$

Each of these equations contain a finite number of terms and the constants $A_0, A_1, A_2, \ldots$ may readily be determined by solving them. In fact we find that any constant A_n may be expressed as

$$A_n = \int_{-\infty}^{+\infty} P_n(x)\,F(x)dx$$

where $P_n(x)$ is a polynomial in x of degree not greater than n. An analysis of the underlying facts reveals that what Charlier has actually done is to show that under the conditions listed in the theorem there exists a uniquely determined set of polynomials $P_0(x), P_1(x), \ldots, P_n(x), \ldots, P_n(x)$ at most of degree n, biorthogonal to the set of derivatives or functions of $f(x)$, i. e. satisfy the biorthogonality conditions:

$$\int_{-\infty}^{+\infty} P_n(x) f^{(m)}(x)dx = 0 \text{ for } m \neq n$$

$$= 1 \text{ for } m = n$$

Further a study of the coefficients of these polynomials shows that

$$\frac{d\,P_n(x)}{dx} = -P_{n-1}(x),$$

i. e. we have the following theorem:

THEOREM: *If* $f(x)$ *satisfy the conditions* (a), (b), (c), *and* (d) *of Charlier's theorem for series* (A) *and if* $P_0(x)$, $P_1(x)$, . . . $P_n(x)$ . . . *is the system of polynomials in* x, $P_n(x)$ *of degree at most* n, *which is biorthogonal to* $f(x)$ *and its derivatives, i. e. satisfies the conditions*

$$\int_{-\infty}^{+\infty} P_n(x)\, f^{(m)}(x)\, dx = 0 \text{ for } m \neq n$$

$$= 1 \text{ for } m = n$$

then

$$\frac{dP_n(x)}{dx} = -P_{n-1}(x)$$

This can readily be shown to be true directly from a use of the biorthogonal property. For integrating by parts we obtain:

$$\int_{-\infty}^{+\infty} P_n(x) f^{(m)}(x)\, dx = P_n(x) f^{(m-1)}(x) \Big|_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} P_n'(x)\, f^{(m-1)}(x)\, dx.$$

The first half of the right hand side of this equation vanishes due to condition (c) of Charlier's theorem for series (A). For the second half we have

$$-\int_{-\infty}^{+\infty} P_n'(x) f^{(m-1)}(x)\, dx = 0 \text{ for } m \neq n$$

$$= 1 \text{ for } m = n$$

But we know that

$$+\int_{-\infty}^{+\infty} P_{n-1}(x)\, f^{(m-1)}(x)\, dx = 0 \text{ for } m \neq n$$

$$= 1 \text{ for } m = n$$

determines uniquely the polynomials $P_{n-1}(x)$. It follows that

$$dP_n(x)/dx = -P_{n-1}(x)$$

A corollary to this last theorem may be stated as follows:

COROLLARY:

$$If \quad \int_{-\infty}^{+\infty} P_n(x)\, f^{(m)}(x)\, dx = 0 \qquad for\ m \neq n$$

$$= a_n \qquad for\ m = n$$

$a_i \neq 0 \qquad (i = 0, 1, 2, \cdots\cdots), then$

$$dP_n(x)/dx = -\frac{a_n}{a_{n-1}} P_{n-1}(x)$$

The proof is similar to the one just given. Integration by parts gives the following result:

$$-\int_{-\infty}^{+\infty} f^{(m-1)}(x)\, P_n'(x)\, dx = 0 \qquad \text{for } m \neq n$$

$$= a_n \qquad \text{for } m = n$$

But we know that

$$\int_{-\infty}^{+\infty} f^{(m-1)}(x)\, P_{n-1}(x)\, dx = 0 \qquad \text{for } m \neq n$$

$$= a_{n-1} \qquad \text{for } m = n$$

Therefore we may conclude that

$$-\frac{1}{a_n} \frac{dP_n(x)}{dx} = \frac{1}{a_{n-1}} P_{n-1}(x)$$

or

$$\frac{dP_n(x)}{dx} = -\frac{a_n}{a_{n-1}} P_{n-1}(x)$$

An illustration of this corollary is the case of the well known Hermite polynomials which are involved in Charlier's first paper.[1] These satisfy the conditions

[1] C. V. L. Charlier, loc. cit. Charlier uses as $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-b)^2}{2\sigma^2}}$. In this paper we shall use the simpler basic function e^{-x^2}.

$$\int_{-\infty}^{+\infty} H_m(x)\, H_n(x)\, e^{-x^2} dx = 0 \qquad \text{for } m \neq n$$

$$= 2^n n! \sqrt{\pi} \qquad \text{for } m = n$$

and

$$H_n(x) e^{-x^2} = (-1)^n d^n (e^{-x^2})/dx^n.$$

Hence

$$\int_{-\infty}^{+\infty} H_m(x)\, d^n(e^{-x^2})/dx^n \; dx = 0 \qquad \text{for } m \neq n$$

$$= (-2)^n n! \sqrt{\pi} \qquad \text{for } m = n$$

If then $f(x) = e^{-x^2}$ and $a_n = (-2)^n n! \sqrt{\pi}$ our corollary applies, i. e. we have

$$d H_n(x)/dx = 2n H_{n-1}(x)$$

We might further observe that if $a_n = (-1)^n n!$ then the polynomials $P_n(x)$ form a system of Appell polynomials[1] satisfying the relation

$$d P_n(x)/dx = n P_{n-1}(x)$$

the n th polynomial being the coefficient of $h^n/n!$ in the expansion of $a(h)\, e^{hx}$ where

$$a(h) = \alpha_0 + \frac{h}{1!}\alpha_1 + \frac{h^2}{2!}\alpha_2 + \cdots + \frac{h^n}{n!}\alpha_n + \cdots$$

The fact that differentiation of the n th polynomial results in the negative of the $(n-1)$th polynomial, shows that the n th polynomial may be obtained by integrating the $(n-1)$ th one,

---

[1] M. P. Appell, "Sur une classe de Polynomes," Annales Scientifiques de L'Ecole Normale Superioure, Vol. IX, series 2 (1880), pp. 119-120.

which will consequently determine all of the terms of the n th polynomial except the constant. This constant may be found from any of the conditions of biorthogonality. The simplest of these conditions is:

$$\int_{-\infty}^{+\infty} P_n(x) f(x)\, dx = 0$$

Setting

$$P_n(x) = -\int_0^x P_{n-1}(x)\, dx + c$$

gives $$\int_{-\infty}^{+\infty} [-\int_0^x P_{n-1}(x)\, dx + c] f(x)\, dx = 0$$

and so $$c = \frac{\int_{-\infty}^{+\infty} [\int_0^x P_{n-1}(x)\, dx] f(x)\, dx}{\int_{-\infty}^{+\infty} f(x) dx}$$

so that $$P_n(x) = -\int_0^x P_{n-1}(x) dx + \frac{\int_{-\infty}^{+\infty} [\int_0^x P_{n-1}(x)\, dx] f(x) dx}{\int_{-\infty}^{+\infty} f(x) dx}$$

This gives a very simple and elegant method of writing down successively the polynomials associated with any function $f(x)$ satisfying the conditions of the theorem.

Using the Charlier notation

$$\lambda_n = \frac{\int_{-\infty}^{\infty} x^n f(x) dx}{n!}$$

and observing that $P_0(x) = 1/\lambda_0$, we obtain the following

polynomials:

$$P_1(x) = -\int_0^x P_0(x)dx + \frac{\int_{-\infty}^{+\infty}\left[\int_0^x P_0(x)dx\right] f(x)dx}{\int_{-\infty}^{+\infty} f(x)dx}$$

$$= -\frac{x}{\lambda_0} + \frac{\lambda_1}{\lambda_0^2},$$

$$P_2(x) = -\int_0^x P_1(x)dx + \frac{\int_{-\infty}^{+\infty}\left[\int_0^x P_1(x)dx\right] f(x)dx}{\int_{-\infty}^{+\infty} f(x)dx}$$

$$= \frac{x^2}{\lfloor 2\,\lambda_0} - \frac{\lambda_1 x}{\lambda_0^2} + \frac{\lambda_1^2}{\lambda_0^3} - \frac{\lambda_2}{\lambda_0^2},$$

$$P_3(x) = -\int_0^x P_2(x)dx + \frac{\int_{-\infty}^{\infty}\left[\int_0^x P_2(x)dx\right] f(x)dx}{\int_{-\infty}^{+\infty} f(x)dx}$$

$$= -\frac{x^3}{\lfloor 3\,\lambda_0} + \frac{\lambda_1 x^2}{\lfloor 2\,\lambda_0^2} - \frac{\lambda_1^2 x}{\lambda_0^3} + \frac{\lambda_2 x}{\lambda_0^2} + \frac{\lambda_1^3}{\lambda_0^4} - \frac{2\lambda_2\lambda_1}{\lambda_0^3} + \frac{\lambda_3}{\lambda_0^2},$$

. . . . . .

. . . . . .

2. Just as the Hermite polynomials, based as they are on the derivatives of e^{-x^2}, are the starting point for expansions of the Gram-Charlier type and for the theorem just considered, so the Laguerre polynomials defined by $d^n(a+bx)^n e^{-x}/dx^n$ suggest an expansion of the type

$$F(x) = C_0 f_0(x)\varphi(x) + C_1 \frac{d}{dx} f_1(x)\varphi(x) + C_2 \frac{d^2}{dx^2} f_2(x)\varphi(x) + \cdots$$

where $f_n(x)$ is a polynomial in x. As a matter of fact we can state the following theorem:

THEOREM: If $\varphi(x)$ *is a function such that*

(1) $\varphi(x)$ *and all its derivatives are continuous for all real values of* x,

(2) $\varphi(x)$ *and its derivatives are zero at* $x = +\infty$ *and* $-\infty$,

(3) $\lim_{x \to \pm\infty} x^n \varphi^m(x) = 0$,

(4) $\{f_n(x)\}$ *is a sequence of polynomials in* x *such that* $\int_{-\infty}^{+\infty} f_n(x)\varphi(x) \neq 0$,

then there exists a unique sequence of polynomials $P_m(x)$, $P_m(x)$ *at most of degree* m, *such that*

$$\int_{-\infty}^{+\infty} P_m(x) \frac{d^n}{dx^n} f_n(x) \varphi(x)\, dx = 0 \qquad \text{for } m \neq n$$

$$= 1 \qquad \text{for } m = n$$

If $f_n(x)$ *is at most of degree* n, *then the determination of* $P_m(x)$ *depends at most upon the moments of* φ *of order* n.[1]

The method of proof is modelled on Charlier's proof for the preceding case. By substituting in the n th integration by parts formula

$$\int u(x) v^{n+1}(x)dx = uv^{(n)} - u'v^{(n-1)} + u''v^{(n-2)} - \ldots + (-1)^n u^{(n)} v + (-1)^{n+1} \int u^{(n+1)}(x) v(x)\, dx,$$

we have

[1]The Laguerre polynomials are not a special case of this because there the interval of integration is $-a/b$ to $+\infty$.

$$\int_{-\infty}^{+\infty} P_m(x)\frac{d^n}{dx^n} f_n(x)\phi(x)dx = \left\{P_m(x)\frac{d^{n-1}}{dx^{n-1}} f_n(x)\phi(x) - \frac{dP_m(x)}{dx}\frac{d^{n-2}}{dx^{n-2}} f_n(x)\phi(x)\right.$$
$$+\frac{d^2}{dx^2}P_m(x)\left[\frac{d^{n-3}}{dx^{n-3}} f_n(x)\phi(x)\right]$$
$$+\cdots+(-1)^{n-1}\left[\frac{d^{n-1}}{dx^{n-1}}P_m(x)\right] f_n\phi(x)\Big\}_{-\infty}^{+\infty}$$
$$+(-1)^n\int_{-\infty}^{+\infty}\left[\frac{d^n}{dx^n}P_m(x)\right] f_n(x)\phi(x)\,dx$$
$$=(-1)^n\int_{-\infty}^{+\infty}\left[\frac{d^n}{dx^n}P_m(x)\right] f_n(x)\phi(x)\,dx$$

because of conditions (2) and (3) on $\varphi(x)$. As a consequence, if $n > m$ then $\frac{d^n}{dx^n} P_m(x) = 0$, so that for $n > m$

$$\int_{-\infty}^{+\infty} P_m(x)\frac{d^n}{dx^n} f_n(x)\phi(x)dx = 0$$

that is to say $P_m(x)$ is orthogonal to $\frac{d^n}{dx^n} f_n(x)\phi(x)$ provided $n > m$ Hence $P_n(x)$ must satisfy only the following $n+1$ equations:

$$\int_{-\infty}^{+\infty} P_m(x)\frac{d^n}{dx^n} f_n(x)\phi(x)dx = 0$$

$$\int_{-\infty}^{+\infty} P_n(x)\frac{d}{dx} f_1(x)\phi(x)dx = (-1)\int_{-\infty}^{+\infty}\frac{dP_n(x)}{dx} f_1(x)\phi(x)dx = 0$$

$$\int_{-\infty}^{+\infty} P_n(x)\frac{d^2}{dx^2} f_2(x)\phi(x)dx = (-1)^2\int_{-\infty}^{+\infty}\frac{d^2P_n(x)}{dx^2} f_2(x)\phi(x)dx = 0$$

. . . . . .

$$\int_{-\infty}^{+\infty} P_n(x)\frac{d^n}{dx^n} f_n(x)\phi(x)dx = (-1)^n\int_{-\infty}^{+\infty}\frac{d^n}{dx^n} P_n(x) f_n(x)\phi(x)dx = 1$$

Replacing now $P_n(x)$ by $a_0 + a_1x + a_2x^2 + \cdots + a_nx^n + \cdots$ gives us the system of algebraic equations to be satisfied by $a_0, a_1, \ldots, a_n$, viz.:

$$a_0\int_{-\infty}^{+\infty} f_0(x)\phi(x)dx + a_1\int_{-\infty}^{+\infty} x f_0(x)\phi(x)dx$$

$$+ a_2\int_{-\infty}^{+\infty} x^2 f_0(x)\phi(x)dx + \cdots\cdots + a_n\int_{-\infty}^{+\infty} x^n f_0(x)\phi(x)dx = 0$$

$$a_1\int_{-\infty}^{+\infty} f_1(x)\phi(x)dx + 2a_2\int_{-\infty}^{+\infty} x f_1(x)\phi(x)dx + \cdots + na_n\int_{-\infty}^{+\infty} x^{n-1} f_1(x)\phi(x)dx = 0$$

$$2a_2\int_{-\infty}^{+\infty} f_2(x)\phi(x)dx + \cdots + n(n-1)a_n\int_{-\infty}^{+\infty} x^{n-2} f_2(x)\phi(x)dx = 0$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$(n-2)!a_{n-2}\int_{-\infty}^{+\infty} f_{n-2}(x)\phi(x)dx + \frac{(n-1)!}{1!}a_{n-1}\int_{-\infty}^{+\infty} x f_{n-2}(x)\phi(x)dx$$

$$+ \frac{n!}{2!}a_n\int_{-\infty}^{+\infty} x^2 f_{n-2}(x)\phi(x)dx = 0$$

$$(n-1)!a_{n-1}\int_{-\infty}^{+\infty} f_{n-1}(x)\phi(x)dx + \frac{n!}{1!}a_n\int_{-\infty}^{+\infty} x f_{n-1}(x)\phi(x)dx = 0$$

$$(-1)^n n!a_n\int_{-\infty}^{+\infty} f_n(x)\phi(x)dx = 1$$

We have here a unique determination of a_n if the determinant of the coefficients is $\neq 0$. This is true since the determinant $\Delta = (-1)^n(\int f_0\phi)(\int f_1\phi)\ldots.(\int f_n\phi)$ is $\neq 0$ because of the condition (4) on ϕ. If $f_n(x)$ is at most of degree n, it is obvious that the determination of the $P_n(x)$ resulting from the coefficients a_n depends at most upon the moments of ϕ of order n.

The first three polynomials of the type considered in the last theorem have the following form, the limits of integration being $-\infty$ and $+\infty$ in each case.

$$P_1(x) = \frac{\int x\phi(x)dx}{\int f_1(x)\phi(x)dx\int\phi(x)dx} - \frac{x}{\int f_1(x)\phi(x)dx}$$

$$= \frac{1}{\int f_1(x)\phi(x)dx}\left[\frac{\int x\phi(x)dx}{\int\phi(x)dx} - x\right],$$

$$P_2(x) = \frac{\int x f_1(x)\phi(x)dx\int x\phi(x)dx}{\int f_2(x)\phi(x)dx\int f_1(x)\phi(x)dx\int\phi(x)dx} - \frac{\int x^2\phi(x)dx}{2!\int f_2(x)\phi(x)dx\int\phi(x)dx}$$

$$- \frac{x \int x f_1(x) \phi(x) dx}{\int f_2(x) \phi(x) dx \int f_2(x) \phi(x) dx} + \frac{x^2}{2! \int f_2(x) \phi(x) dx},$$

$$P_3(x) = \frac{\int x f_2(x) \phi(x) dx \int x f_1(x) \phi(x) dx \int x \phi(x) dx}{\int f_3(x) \phi(x) dx \int f_2(x) \phi(x) dx \int f_1(x) \phi(x) dx \int \phi(x) dx}$$

$$- \frac{\int x^2 f_1(x) \phi(x) dx \int x \phi(x) dx}{2! \int f_3(x) \phi(x) dx \int f_1(x) \phi(x) dx \int \phi(x) dx}$$

$$- \frac{\int x f_2(x) \phi(x) dx \int x^2 \phi(x) dx}{2! \int f_3(x) \phi(x) dx \int f_2(x) \phi(x) dx \int \phi(x) dx}$$

$$+ \frac{\int x^3 \phi(x) dx}{3! \int f_3(x) \phi(x) dx \int \phi(x) dx} - \frac{x \int x f_2(x) \phi(x) dx \int x f_1(x) \phi(x) dx}{\int f_3(x) \phi(x) dx \int f_2(x) \phi(x) dx \int f_1(x) \phi(x) dx}$$

$$+ \frac{x \int x^2 f_1(x) \phi(x) dx}{2! \int f_3(x) \phi(x) dx \int f_1(x) \phi(x) dx} + \frac{x^2 \int x f_2(x) \phi(x) dx}{2! \int f_3(x) \phi(x) dx \int f_2(x) \phi(x) dx}$$

$$- \frac{x^3}{3! \int f_3(x) \phi(x) dx}.$$

CHAPTER II

Polynomials Connected with Pearson's Differential Equation

1. In the work in mathematical statistics a large number of the problems that require study involve data properly classified into groups and about which further information is sought. This data is often classified to form a frequency distribution. The frequency distribution when grouped may appear to lie on a certain curve. If it can be shown that this curve is a mathematical curve, i. e. one for which we are able to set up an equation, then this frequency distribution can be readily examined and studied.

There are very few frequency distributions which actually conform to known mathematical equations. However, there are certain curves which seem to lend themselves much better to statistical manipulations than others. Among the most commonly used of these are the so-called Pearson type curves. Pearson[1] showed in a series of three articles how he obtained the equations of twelve distinct curves and this was done by considering the differential equation

$$\frac{1}{y}\frac{dy}{dx} = \frac{a_0 + a_1 x}{b_0 + b_1 x + b_2 x^2}$$

and solving it, after assigning particular values to the parameters a_0 , a_1 , b_0 , b_1 , and b_2 . The equations of these curves and the differential equations from which they were derived are as follows:

[1]Karl Pearson, loc. cit.

Type	Equation	Differential Equation
I	$y = y_0 \left(1+\frac{x}{a}\right)^{\nu a} \left(1-\frac{x}{b}\right)^{\nu b}$	$\frac{dy}{dx} = \frac{\nu(a+b)x}{(a+x)(b-x)}\, y$
II	$y = y_0 \left(1-\frac{x^2}{a^2}\right)^{m}$	$\frac{dy}{dx} = \frac{-2mx}{a^2-x^2}\, y$
III	$y = y_0 \left(1+\frac{x}{a}\right)^{\nu a} e^{-\nu x}$	$\frac{dy}{dx} = \frac{-\nu x}{a+x} y$
IV	$y = y_0 \left(1+\frac{x^2}{a^2}\right)^{-m} e^{-\nu \arctan \frac{x}{a}}$	$\frac{dy}{dx} = \frac{-2mx-\nu a}{a^2+x^2} y$
V	$y = y_0\, x^{-h} e^{-\frac{\nu}{x}}$	$\frac{dy}{dx} = \frac{\nu-hx}{x^2} y$
VI	$y = y_0\, (x-a)^q x^{-h}$	$\frac{dy}{dx} = \frac{ha+(q-h)x}{x^2-\partial x} y$
VII	$y = y_0\, e^{-\frac{x^2}{2\sigma^2}}$	$\frac{dy}{dx} = \frac{-x}{\sigma^2} y$
VIII	$y = y_0 \left(1+\frac{x}{a}\right)^{-m}$	$\frac{dy}{dx} = \frac{-m}{a+x} y$
IX	$y = y_0 \left(1+\frac{x}{a}\right)^{m}$	$\frac{dy}{dx} = \frac{m}{a+x} y$
X	$y = \frac{n}{\sigma} e^{\pm \frac{x}{\sigma}}$	$\frac{dy}{dx} = \pm \frac{1}{\sigma} y$
XI	$y = y_0\, x^{-m}$	$\frac{dy}{dx} = \frac{-m}{x} y$
XII	$y = y_0 \left(\frac{a_1+x}{a_2-x}\right)^{p}$	$\frac{dy}{dx} = \frac{h(a_1+a_2+2x)}{(a_1+x)(a_2+x)} y$

The curves most widely used are the normal curve of error, which Pearson calls Type VII, and the Type III curve.

Suppose a Pearson curve $f(x)$ has been found which seems to fit a given distribution fairly well. The question may well be asked: Is it possible by means of analytic methods to approach even nearer to the given distribution? For example, would it be possible to use this approximate function as the $f(x)$ in the Charlier series (A) and thus obtain a closer approximation to the observed frequency function.

Charlier in his paper "Ueber die Darstellung willkürlicher Functionen"[1] considered this question for $\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-b)^2}{2\sigma^2}}$, i. e. the normal curve of error. He showed that using this $\varphi(x)$ reduced the series (A) to the form:

$$(A') \quad F(x) = a_0\varphi(x) + a_3\varphi^{(3)}(x) + a_4\varphi^{(4)}(x) + \cdots + a_n\varphi^{(n)}(x) + \cdots$$

the first and second derivative terms vanishing due to the proper choice of constants. This series (A′) is frequently referred to as the Gram-Charlier Type A series. It is worthwhile to note that this $\varphi(x)$ is the same one whose derivatives we found in the first chapter resulted in the Hermite polynomials. These polynomials have the following interesting properties[2]:

(1) $dH_n(x)/dx = 2nH_{n-1}(x)$

(2) $H_{n+1}(x) - 2xH_n(x) + 2nH_{n-1}(x) = 0$

(3) $H_n''(x) - 2xH_n'(x) + 2nH_n(x) = 0$

The first of these relations shows that the derivative of any Hermite polynomial corresponds to the preceding polynomial multi-

[1] C. V. L. Charlier, loc. cit.

[2] R. Courant and D. Hilbert, Methoden der Mathematischen Physik, 1, pp. 76.

plied by $2n$. The second equation is a recurrence relation between the $(n+1)$ th, n th and $(n-1)$ th polynomials, while the third relation is a differential equation of the second order involving only the n th polynomial.

The use of the equations of the other Pearson type curves as the $f(x)$ in the original Charlier series has in recent years been studied by Romanovsky. In the first[1] of two articles, he discusses the Pearson Type I, II and III curves as well as the Type VII—the normal curve referred to in the last paragraph. Just as the normal curve of error requires the use of the Hermite polynomials, he found that the Type I curve and Type II, which is a special case of Type I, involved the Jacobi polynomials

$$G_n(h,q,x)=\frac{x^{1-q}(1-x)^{q-h}}{q(q+1)\cdots(q+n-1)}\frac{d^n}{dx^n}\left[x^{q+n-1}(1-x)^{h+n-q}\right].$$

The n 'th Jacobi polynomial satisfies the second order differential equation[2].

$$x(1-x)G_n''(x)+[q-(p+1)x]G_n'(x)+(p+n)nG_n(x)=0$$

which corresponds to property (3) mentioned for the Hermite polynomials above. The Type III curve involves the Laguerre polynomials[3] defined by

$$L_n(x)=e^x\frac{d^n}{dx^n}(x^n e^{-x})$$

and these in turn satisfy the recurrence relation

---

[1] V. Romanovsky: "Generalization of some types of the frequency curves of Professor Pearson." op. at pp. 106-117.

[2] R. Courant and D. Hilbert, op. cit., Vol. I, p. 75.

[3] R. Courant and D. Hilbert, op. cit., pp. 77-78.

$$L_{n+1}(x)-(2n+1-x)L_n(x)+n^2L_{n-1}(x)=0$$

and the differential equation

$$L'_n(x) - n L'_{n-1}(x) = -n L_{n-1}(x)$$

In the second article[2], Romanovsky reviews the cases of the Type IV, V and VI curves. The generalization of the Type IV curve gives the polynomial

$$P_n(m,x)=(a^2+x^2)^m e^{\nu\theta}\frac{d^n}{dx^n}\left[(a^2+x^2)^{-m+n}e^{-\nu\theta}\right]$$

where $\theta = \arctan x/a$. These polynomials possess properties similar to the other polynomials mentioned, viz.:

$$P_{n+1}(n+1,x)=\left[2(n+1-m)x-\nu a\right]P_n(n,x)$$

$$+2n\left[n+1-m\right](a^2+x^2)P_{n-1}(n,x)$$

and

$$(a^2+x^2)P''_n(n,x)+\left[2(1-m)x-\nu a\right]$$

$$P'_n(n,x)-n(n+1-2m)P_n(n,x)=0$$

Similarly for the Type V curve he finds the polynomials

$$P_n(h,x)=x^h e^{\frac{\nu}{x}}\frac{d^n}{dx^n}\left(x^{-h+2n}e^{-\frac{\nu}{x}}\right).$$

Also the relations

---

[2]V. Romanovsky, "Sur quelques Classes nouvells de Polynomes othogonaux," loc. cit.

$$P_{n+1}(n+1)x) = \left[(2n+2-p)x+\gamma\right]P_n(n,x) + n(2n+2-p)x^2P_{n-1}(n,x)$$

and

$$x^2P_n''(n,x) + \left[x(2-p)+\gamma\right]P_n'(n,x) - n(n+1-p)\,P_n(n,x) = 0$$

hold.

Finally for the Type VI curve Romanovsky gets the polynomials:

$$P_n(-h,q,x) = (x-a)^{-q}x^h\frac{d^n}{dx^n}\left[(x-a)^{q+n}x^{-h+n}\right]$$

and the relations:

$$P_{n+1}(n+1,x) = \left[(-p+1)(x-a)+(q+1)x\right]P_n(n,x) + x(x-a)P_n'(n,x),$$

$$x(x-a)P_n''(n,x) + \left[(-p+1)(x-a)+(q+1)x\right]P_n'(n,x) - n(n+1+q-p)P_n(n,x) = 0.$$

We note, therefore, that if a solution of the Pearson differential equation is used as the generating function $f(x)$ in the Gram-Charlier series, that a distinct set of polynomials results in each case and that these polynomials satisfy certain recurrence relations and differential equations. These properties are not found in the case of functions such as sech x and sech $^m x$, which were discussed as generating functions by Charlier[1] and by Roa[2] respectively. The successive derivatives of the

---

[1]C. V. L. Charlier, "Ueber die Darstellung willkürlicher Funktionen," loc. cit., pp. 18-22.

[2]Emeterio Roa, "A Number of new generating Functions with Applications to Statistics," Doctor's Thesis, University of Michigan, 1923.

sech x do not result in polynomials such as the Hermite or Jacobi ones.

Since the generalization of the solutions of the Pearson curves leads to distinct sets of polynomials and since these polynomials satisfy certain fundamental relations, we are led to inquire whether these polynomials are not special cases of a general polynomial and may be obtained from it by specializing the coefficients and further whether such general polynomials, if they do exist, will satisfy certain recurrence relations and differential equations. These problems are among those which we shall consider in this chapter.

2. In order that we may develop the generalized polynomials, let us consider the Pearson differential equation where the numerator is of the first and the denominator of the second degree, i. e.

$$\frac{1}{y}\frac{dy}{dx} = \frac{a_0 + a_1 x}{b_0 + b_1 x + b_2 x^2}$$

For convenience we shall denote the numerator by N and the denominator by D. We then have the following theorem:

THEOREM: *If y is a non-identically zero solution of*

$$\frac{dy}{dx} = \frac{N}{D} y \tag{1}$$

then $\frac{D^n}{y}\frac{d^n y}{dx^n}$ is a polynomial of degree at most n.

The proof will proceed by mathematical induction. It is obvious that the theorem holds for $n = 1$, $P_1(x)$ being N. Since it is true that

$$D\frac{dy}{dx} = Ny$$

we obtain by differentiation

$$D\frac{d^2y}{dx^2} + D'\frac{dy}{dx} = N\frac{dy}{dx} + N'y ,$$

or using (1) and multiplying the equation through by D we get

$$D^2 \frac{d^2y}{dx^2} = (N^2 - ND' + N'D)y$$

Since D' is linear and N' is a constant, it is obvious that $(N^2 - ND + N'D)$ is at most of degree 2.

Assume then that the statement holds for $m \leq n$ and we have

$$(2) \qquad D^n \frac{d^n y}{dx^n} = P_n(x)\, y \,.$$

Differentiation gives

$$nD^{n-1}D' \frac{d^n y}{dx^n} + D^n \frac{d^{n+1}}{dx^{n+1}} = P_n(x)\frac{dy}{dx} + \frac{dP_n(x)}{dx} y.$$

Multiplying through by D we get

$$D^{n+1} \frac{d^{n+1}}{dx^{n+1}} = DP_n(x)\frac{dy}{dx} - nD^n D' \frac{d^n y}{dx^n} + D\frac{dP_n(x)}{dx} y,$$

and using (1) and (2), we have

$$P_{n+1}(x)\cdot y = NP_n(x)y - nD'P_n(x)y + D\frac{dP_n(x)}{dx} y$$
$$= \left[NP_n(x) - nD'P_n(x) + D\frac{dP_n(x)}{dx}\right] y.$$

The coefficient of y is obviously a polynomial of degree at most $n+1$. Incidentally we have derived the relation:

$$(I) \qquad P_{n+1}(x) = P_n(x)(N - nD') + D\frac{dP_n(x)}{dx}$$

an equation which gives the $(n+1)$ th polynomial in terms of the n th polynomial and its first derivative $P_n'(x)$.

3. More generally we have:

THEOREM: *If y is a non-identically zero solution of* (1), *then*

$$\frac{1}{y} D^{n-k} \frac{d^n}{dx^n} D^k y$$

is a polynomial $P_n(k, x)$, $P_n(k, x)$ *is at most of degree* n *in* x. *In particular if* $k = n$, *we have that*

$$\frac{1}{y}\frac{d^n}{dx^n} D^n y$$

is a polynomial in x *of degree at most* n.

This theorem can be proved directly following the lines of the preceding theorem, but it is simpler to obtain it as an immediate consequence of this theorem and the following lemma:

LEMMA: *If* y *satisfy the differential equation* (1) *then* $D^k y$, *where* k *is any real number, satisfies a differential equation of the same type, viz.:*

$$\frac{d}{dx}(D^k y) = \frac{N + kD'}{D} D^k y$$

Let $u = D^k y$

Then logarithmic differentiation gives at once

$$\frac{1}{u}\frac{du}{dx} = K\frac{D'}{D} + \frac{1}{y}\frac{dy}{dx} = \frac{N + kD'}{D}$$

It follows from this lemma that any result which we derive concerning the polynomials $P_n(x) = \frac{1}{y} D^n \frac{d^n y}{dx^n}$ where y satisfies $D\, dy/dx = Ny$, is immediately extensible to the polynomials $P_n(k, x) = \frac{1}{y} D^{n-k} \frac{d^n}{dx^n} D^k y$ by replacing N by $N + kD'$. In particular relation (I) becomes

$$(I_k)\quad P_{n+1}(k+1, x) = \left[N + (k-n+1)D'\right] P_n(k+1, x) + D \cdot \frac{d P_n(k+1, x)}{dx}$$

which for $k = n$ reduces to

$$(I_n)\quad P_{n+1}(n+1, x) = (N + D') P_n(n+1, x) + D \frac{d P_n(n+1, x)}{dx}.$$

We single out the case $k = n$ because of the fact that this case parallels most closely the Charlier or Hermite polynomial case. For in this latter case the n'th derivative of the generating function e^{-x^2} is the product of the generating function and a polynomial of degree n. *So in the case of any solution* y *of a Pearson differential equation, the* n *th derivative of* $D^n y$ *is the product of the generating function* y *and a polynomial of degree at most* n.

By means of relation (I), we can write down the successive polynomials $P_1(x)$, $P_2(x)$, . . . The first five polynomials may be written as follows:

$$P_1(x) = N,$$

$$P_2(x) = (N - D')P_1(x) + D\frac{dP_1(x)}{dx} = N^2 - ND' + N'D,$$

$$P_3(x) = (N - 2D')P_2(x) + D \cdot \frac{dP_2(x)}{dx}$$

$$= N^3 - 3N^2D' + 3NN'D + 2ND'^2 - 2N'D'D - NDD'',$$

$$P_4(x) = (N - 3D')P_3(x) + D\frac{dP_3(x)}{dx}$$

$$= N^4 - 6N^3D' + 6N^2N'D + 11N^2D'^2 - 14NN'DD' - 4N^2DD'',$$

$$= -6ND'^3 + 6N'DD'^2 + 6NDD'D'' + 3N'^2D^2 - 3N'D^2D'',$$

$$P_5(x) = (N - 4D')P_4(x) + D\frac{dP_4(x)}{dx}$$

$$= N^5 - 10N^4D' + 10N^3N'D + 35N^3D'^2 - 50N^2N'DD'$$

$$- 10N^3DD'' - 50N^2D'^3 + 70NN'DD'^2 - 40N^2DD'D''$$

$$+ 15NN'^2D^2 - 25NN'D^2D'' + 24ND'^4 - 24N'DD'^3$$

$$- 36NDD'^2D'' - 20N'^2D^2D' + 24N'D^2D'D'' + 6ND^2D''^2$$

4. Following the analogy with Hermite polynomials, we obtain next a recurrence relation involving the $(n+1)$ th, n'th and $(n-1)$ th polynomials.

Starting with the original differential equation

$$D\frac{dy}{dx} = Ny$$

we take the n th derivative of both sides, which by Leibnitz's theorem on the derivative of a product gives us, since $\frac{d^3D}{dx^3} = 0$,

$$D\frac{d^{n+1}y}{dx^{n+1}} + nD'\frac{d^ny}{dx^n} + \frac{n(n-1)}{2!}D''\frac{d^{n-1}y}{dx^{n-1}} = N\frac{d^ny}{dx^n} + nN'\frac{d^{n-1}y}{dx^{n-1}}$$

Multiplying this last expression by D^n and collecting terms, we get:

$$D^{n+1}\frac{d^{n+1}y}{dx^{n+1}} + D^n(nD'-N)\frac{d^ny}{dx^n} + D^n\left[\frac{n(n-1)}{2!}D'' - nN'\right]\frac{d^{n-1}y}{dx^{n-1}} = 0$$

Replacing now $D^n\frac{d^ny}{dx^n}$ by $P_n(x)y$ and dividing through by y, we get the recurrence relation

$$\text{(II)}\quad P_{n+1}(x) + (nD'-N)P_n(x) + n\left[\frac{(n-1)}{2!}D'' - N'\right]DP_{n-1}(x) = 0$$

We note that the coefficients of $P_{n+1}(x)$ and $P_n(x)$ are the same as in relation (I) which we found to be

$$P_{n+1}(x) + P_n(x)(nD'-N) = D\frac{dP_n(x)}{dx}.$$

Hence

$$\text{(III)}\qquad \frac{dP_n(x)}{dx} = n\left[N' - \frac{(n-1)}{2}D''\right]P_{n-1}(x)$$

or replacing n by $n+1$ we write:

$$\frac{d P_{n+1}(x)}{dx} = (n+1)(N' - \frac{n}{2} D'') P_n(x) = (n+1)(a_1 - n b_2) P_n(x).$$

This equation is the generalized form of the one for Hermite polynomials, viz.:

$$\frac{d H_n(x)}{dx} = 2n H_{n-1}(x)$$

5. Relations (I) and (III) may now be used to obtain a second order differential equation. Differentiating (I), we get:

$$P'_{n+1}(x) + (nD'' - N') P_n(x) + (nD' - N) P'_n(x)$$
$$- D' P'_n(x) - D P''_n(x) = 0.$$

Substitution of the value $d P_{n+1}(x)/dx$ from (III) gives us:

$$\text{(IV)} \qquad \begin{aligned} & D P''_n(x) + \left[N - (n-1) D'\right] P'_n(x) \\ & - n \left[N' - \frac{(n-1) D''}{2}\right] P_n(x) = 0 \end{aligned}$$

We readily see that the relation found for the Hermite polynomials

$$H''_n(x) - 2x H'_n(x) + 2n H_n(x) = 0$$

is a special case of (IV).

Using the lemma previously proved and replacing N by $N + kD'$ we can write (IV) for the polynomials $P_n(k, x)$ and $P_n(n, x)$:

$$\text{(IV}_k\text{)} \qquad \begin{aligned} & D P''_n(k, x) + \left[N - (n-k-1) D'\right] P'_n(k, x) \\ & - n \left[N' - \frac{(n-2k-1)}{2} D''\right] P_n(k, x) = 0, \end{aligned}$$

$$(IV_n)^1 \qquad \begin{aligned} & DP_n''(n,x)+(N+D')P_n'(n,x) \\ & -n\left[N'+\frac{(n+1)}{2}D''\right]P_n(n,x)=0 \end{aligned}$$

We recognize the second order differential equations mentioned earlier in this chapter for the polynomials of the Pearson Type

---

[1]Since D is any expression of the second degree and N is any expression of the first degree, it is obvious that $P_n(x)$ satisfies a linear equation of the second order of the form:

$$(A_0+A_1x+A_2x^2)y''+(B_0+B_1x)y'+Cy=0$$

where $C=-n\left[(n-1)A_2+B_1\right]$. It may be shown that if a differential equation of the form considered has as one solution a polynomial of degree n then C must be of the form specified. For suppose $Q_n(x)$ satisfies the above differential equation for y. Taking the n'th derivative of this equation we get

$$\frac{n(n-1)}{2!}\cdot 2A_2(n!a_0)+nB_1(n!a_0)+C(n!a_0)=0$$

and solving for C that:

$$C=-n\left[(n-1)A_2+B_1\right].$$

It follows from our work that if a differential equation has the form

$$\begin{aligned} & (A_0+A_1x+A_2x^2)y''+(B_0+B_1x)y' \\ & -n\left[(n-1)A_2+B_1\right]y=0 \end{aligned}$$

then one solution of this differential equation is a polynomial of degree at most n obtained by finding the solution y of the Pearson differential equation

$$\frac{dy}{dx}=\frac{B_0+B_1x-(A_1+2A_2x)}{A_0+A_1x+A_2x^2}y$$

and determining the polynomial

$$P_n(n,x)=\frac{1}{y}\frac{d^n}{dx^n}\left\{\left[A_0+A_1x+A_2x^2\right]^n y\right\}.$$

IV, V and VI as well as the Jacobi and Laguerre polynomials as special cases of formula (IV_n). Some further illustrations of (IV_n) are the Tschebycheff[1] and Legendre[2] polynomials. The Tschebycheff polynomials are developed from the differential equation

$$\frac{dy}{dx} = \frac{x}{1-x^2}\, y$$

and in this case formula (IV_n) becomes:

$$(1-x^2)P_n''(n,x) - xP_n'(n,x) + n^2 P_n(n,x) = 0$$

The Legendre polynomials

$$P_n(x) = \frac{1}{2^n n!}\,\frac{d^n (x^2-1)^n}{dx^n}$$

have as a corresponding differential equation

$$\frac{dy}{dx} = \frac{0 \cdot y}{x^2-1}$$

and in turn formula (IV_n) is written:

$$(x^2-1)P_n''(n,x) + 2xP_n'(n,x) - n(n+1)P_n(n,x) = 0$$

6. Just as in formula (II) we established a recurrence relation for the polynomials $P_n(x)$, let us now obtain one for the polynomials $P_n(n,x)$.

Consider once more the first derivative of $D^k y$, i. e.

$$\frac{d}{dx}(D^{k+1}y) = (K+1)D' \cdot D^k y + D^{k+1} y'$$

$$= \left[N + (K+1)D'\right] D^k y$$

---

[1]R. Courant and D. Hilbert, op. cit., pp. 73-74.

[2]Ibid, pp. 66-69.

Taking the n th derivative of both sides of the equation we get:

$$\frac{d^{n+1}}{dx^{n+1}}(D^{k+1}y) = \left[N+(K+1)D'\right]\frac{d^n}{dx^n}D^k y + n\left[N'+(K+1)D''\right]\frac{d^{n-1}}{dx^{n-1}}D^k y .$$

Multiplying both sides of the equation by D^{n-k} and replacing $D^{n-k}\ \frac{d^n}{dx^n}D^k y$ by $P_n(k,x)y$, we have

$$(V_k)\qquad P_{n+1}(K+1,x) = \left[N+(K+1)D'\right]P_n(K,x) + n\left[N'+(K+1)D''\right]D\cdot P_{n-1}(K,x).$$

In case we set $k = n$, we may write

$$(V_n)\qquad P_{n+1}(n+1,x) = \left[N+(n+1)D'\right]P_n(n,x) + n\left[N'+(n+1)D''\right]D\cdot P_{n-1}(n,x),$$

a recurrence relation similar to (II) and involving the polynomials $P_{n+1}(n+1,x)$, $P_n(n,x)$ and $P_{n-1}(n,x)$.

7. Formula (V_n) may be written in still another form corresponding to formula (I), i. e. a relation consisting of the same terms as (V_n) except that the $(n-1)$ th polynomial $P_{n-1}(n,x)$ is replaced by the first derivative of the n th polynomial $P_n(n,x)$.

In order to obtain this relation we return to formula (III),

$$\frac{dP_n(x)}{dx} = n\left[N'-\frac{(n-1)}{2}D''\right]P_{n-1}(x)$$

and substitute for N the value $N + kD'$ and obtain

$$(III_n)\qquad \frac{dP_n(n,x)}{dx} = n\left[N'+\frac{(n+1)}{2}D''\right]P_{n-1}(n,x)$$

or

$$P_{n-1}(n,x) = \frac{1}{n\left[N' + \left(\frac{n+1}{2}\right)D''\right]} \frac{dP_n(n,x)}{dx}$$

Substituting the value for $P_{n-1}(n,x)$ we thus obtain:

$$\text{(VI)} \qquad P_{n+1}(n+1,x) = \left[N + (n+1)D'\right]P_n(n,x) + \frac{N' + (n+1)D''}{N' + \left(\frac{n+1}{2}\right)D''} \cdot D \cdot \frac{P_n(n,x)}{dx}$$

From symmetry we might expect the fractional coefficient of the derivative $P_n'(n,x)$ to be unity, but unfortunately this is not the case.

8. In looking over the relations existing for the Laguerre polynomials we find one consisting of the first derivatives of the n th and $(n-1)$ th polynomials, and the $(n-1)$ th polynomial,[1] i. e.

$$P_n'(n,x) - nP_{n-1}'(n,x) = -nP_{n-1}(n-1,x)$$

This relation is a special case of another form of formula (VI) which we obtain in the following manner:
Differentiation of (VI) gives us:

$$\frac{dP_{n+1}(n+1,x)}{dx} = \left[N' + (n+1)D''\right]P_n(n,x) + \left[N + (n+1)D'\right]\frac{dP_n(n,x)}{dx}$$

$$+ \frac{N' + (n+1)D''}{N' + \frac{(n+1)}{2}D''} \cdot D \cdot \frac{dP_n(n,x)}{dx} + \frac{N' + (n+1)D''}{N' + \frac{(n+1)}{2}D''} \cdot D \cdot \frac{d^2P_n(n,x)}{dx^2}$$

Substituting the value for $d^2P_n(n,x)/dx^2$ found in (V) changes this last expression to the form:

---

[1] R. Courant and D. Hilbert, op. cit., pp. 77-79.

$$\frac{dP_{n+1}(n+1,x)}{dx} = \left[N' + (n+1)D''\right] P_n(n,x) + \left[N + (n+1)D'\right] \frac{dP_n(n,x)}{dx}$$

$$+ \frac{N' + (n+1)D''}{N' + \frac{(n+1)}{2} D''} \cdot D' \cdot \frac{dP_n(n,x)}{dx} + \frac{N' + (n+1)D''}{N' + \left(\frac{n+1}{2}\right)D''}$$

$$\left[-(N+D')\frac{dP_n(n,x)}{dx} + n\left(\frac{n+1}{2}D'' + N'\right)P_n(n,x)\right],$$

which reduces to

$$\frac{dP_{n+1}(n+1,x)}{dx} = (n+1)\left[N' + (n+1)D''\right] P_n(n,x)$$

(VII)

$$+ \left\{\left[N + (n+1)D'\right] - \frac{N' + (n+1)D''}{N' + \frac{(n+1)}{2} D''} \cdot N\right\} \frac{dP_n(n,x)}{dx} .$$

The special equation mentioned for the Laguerre polynomials will be recognized as a special case of formula (VII) if we recall that for the Laguerre polynomials the differential equation is of the form

$$\frac{dy}{dx} = \frac{p - x}{x} y$$

Substitution of x for D and $(p - x)$ for N reduces (VII) to

$$P'_{n+1}(n+1,x) = -(n+1)P_n(n,x) + (n+1)P'_n(n,x) .$$

9. In this chapter we have defined two general types of polynomials $P_n(x) = \frac{D^n}{y} \frac{d^n y}{dx^n}$

and $P_n(k,x) = \frac{D^{n-k}}{y} \frac{d^n}{dx^n} D^k y$.

The relationships for these polynomials $P_n(x)$ and $P_n(k,x)$ were derived without using the form of the solution of the differential equation. Two fundamental formulas were derived, for $P_n(x)$:

(I) $$P_{n+1}(x) = (N - nD')P_n(x) + D \frac{dP_n(x)}{dx}$$

and for $P_n(n, x)$ the corresponding formula:

$$\text{(VI)}\quad P_{n+1}(n+1,x) = \left[N+(n+1)D'\right]P_n(n,x) + \frac{N'+(n+1)D''}{N'+\frac{(n+1)}{2}D''}D\frac{dP_n(n,x)}{dx}.$$

Two successive polynomials were shown to be related by the relations, for $P_n(x)$:

$$\text{(III)}\qquad \frac{dP_n(x)}{dx} = n\left[N' - \frac{(n-1)}{2}D''\right]P_{n-1}(x)$$

and for $P_n(n, x)$:

$$\text{(III}_n\text{)}\qquad \frac{dP_n(n,x)}{dx} = n\left[N' + \frac{(n+1)}{2}D''\right]P_{n-1}(n,x)$$

In addition we found that it was possible to set up recurrence relations involving the $(n+1)$ th, n th and $(n-1)$ th polynomials and found these to be, for $P_n(x)$:

$$\text{(II)}\quad P_{n+1}(x) + (nD' - N)P_n(x) + n\left[\frac{n-1}{2!}D'' - N'\right]D\cdot P_{n-1}(x) = 0$$

and for $P_n(n, x)$:

$$\text{(V}_n\text{)}\quad P_{n+1}(n+1,x) = \left[N+(n+1)D'\right]P_n(n,x) + n\left[N'+(n+1)D''\right]\cdot D\,P_{n-1}(n,x)$$

We further succeeded in developing a second order differential equation for the n'th polynomial $P_n(x)$:

$$\text{(IV)}\quad DP_n''(x) + \left[N - (n-1)D'\right]P_n'(x) - n\left[N' - \frac{(n-1)}{2}D''\right]P_n(x) = 0$$

and for $P_n(n, x)$:

$$\text{(IV}_n\text{)}\quad DP_n''(n,x) + (N+D')P_n'(n,x) - n\left[N' + \frac{(n+1)}{2}D''\right]P_n(n,x) = 0$$

We also showed that we could derive a relation between the derivatives of the polynomials $P_{n+1}(n+1, x)$, $P_n(n, x)$ and the polynomial $P_n(n, x)$:

$$\text{(VII)} \quad \frac{d P_{n+1}(n+1,x)}{dx} = (n+1)\left[N' + (n+1)D''\right] P_n(n,x) +$$

$$\left\{\left[N + (n+1)D'\right] - \frac{N' + (n+1)D''}{N' + \left(\frac{n+1}{2}\right)D''} \cdot N\right\} \frac{d P_n(n,x)}{dx}.$$

Finally, we noted that all of these formulas and relations apply to the Hermite, Jacobi, Tschebycheff and Legendre polynomials as well as the polynomials derived for the Pearson Type IV, V and VI curves by Romanovsky.

CHAPTER III

1. So far the discussion in this paper has been limited to the treatment of the Gram-Charlier series where the constants $A_0, A_1, A_2, \ldots, A_n, \ldots$ depend upon polynomials in x which are independent of the function $F(x)$, and the generating function $f(x)$ is a solution of the Pearson differential equation, the functions $F(x)$ and $f(x)$ being defined as continuous functions. The work in mathematical statistics involves not only the use of the continuous variate and the continuous function but also the case of the discrete variate and the discontinuous function where this function is defined for equally spaced values.

In dealing with the continuous variate we make use of the theory of the differential and integral calculus, or the calculus of limits, as it is sometimes called. On the other hand, for the discrete variate we turn to the theory of the calculus of finite differences. Further, it usually happens that there exists a parallelism between results based on the derivative and integral and those based on the finite differences and summations. As a consequence, it seems natural to attempt to derive results for the finite difference case paralleling those contained in the first half of this paper. The second part of this paper is devoted to this purpose. The first of the two following chapters considers matters pertaining to Charlier's Type B series which is the finite difference parallel to the Type A series, while the next chapter is devoted to the polynomials connected with the finite difference parallel of the Pearson differential equation.

Charlier in the second half of his article[1] "Ueber die Dar-

*C. V. L. Charlier, op. cit., pp. 23-35.

stellung willkürlicher Funktionen" considers a real valued function $F(x)$ and asserts that it may be formally expanded in terms of another function and its successive differences. Stated as a theorem, this may be written as follows:

CHARLIER'S THEOREM FOR SERIES B: *Any real valued function $F(x)$ which vanishes for $x = \infty$ and $-\infty$, may be formally expanded in terms of another function $g(x)$ and its successive differences in the form*

$$(B)\quad F(x) = B_0 g(x) + B_1 \Delta g(x) + B_2 \Delta^2 g(x) + \cdots + B_n \Delta^n g(x) + \cdots$$

where $g(x)$ possesses the properties:

(a) *$g(x)$ and its differences are defined for all real values of x,*

(b) *$g(x)$ and its differences vanish for $x = +\infty$ and $-\infty$,*

(c) $x^m \Delta^n g(x) \Big|_{-\infty}^{+\infty} = 0$ *for all real values of m and n.*

(d) $\Delta^{-1} g(x) \Big|_{-\infty}^{+\infty} \neq 0.$

Paralleling the theory of the first half of his paper, Charlier determines the constants $B_0, B_1, B_2, \ldots, B_n, \ldots$ and finds that they may be expressed by the equation

$$B_n = \sum_{-\infty}^{+\infty} Q_n(x)\, F(x) = \Delta^{-1} Q_n(x)\, F(x) \Big|_{-\infty}^{+\infty}$$

where $Q_n(x)$ is a polynomial in x of degree not greater than n. Analyzing the answers that he obtains for $Q_n(x)$, we find that these polynomials form a uniquely determined set of polynomials $Q_0(x), Q_1(x), \ldots, Q_n(x), \ldots, Q_n(x)$ at most of degree n, biorthogonal in the sum sense to the successive differences of the function $g(x)$, i. e. they satisfy the biorthogonality conditions for the inverse of differences:

$$\Delta^{-1} Q_n(x) \Delta^m g(x) \Big|_{-\infty}^{+\infty} = \begin{cases} 0 & \text{for } n \neq m \\ 1 & \text{for } n = m. \end{cases}$$

Charlier does not observe that the polynomials $Q_n(x)$ bear a definite relation to one another, i. e.

$$\Delta Q_n(x) = -Q_{n-1}(x+1),$$

a relation similar to the one found for the polynomials $P_n(x)$ in Chapter I. We may state these facts in the following theorem:

THEOREM: *If $g(x)$ satisfy the conditions* (a), (b), (c), *and* (d) *of Charlier's Theorem for series* B *and if $Q_0(x)$, $Q_1(x)$, . . ., $Q_n(x)$, . . . is the system of polynomials in x, $Q_n(x)$ of degree at most n, which is biorthogonal to $f(x)$ and its differences, i. e. satisfies the conditions*

$$\Delta^{-1} Q_n(x) \Delta^m g(x) \Big|_{-\infty}^{+\infty} = \begin{cases} 0 \text{ for } n \neq m \\ 1 \text{ for } n = m \end{cases}$$

then

$$\Delta Q_n(x) = -Q_{n-1}(x+1).$$

The proof requires the use of the finite integration by parts formula:

$$\Delta^{-1} u_x v_x = u_x \Delta^{-1} v_x - \Delta^{-1} \left[\Delta u_x \cdot \Delta^{-1} v_{x+1} \right].$$

Applying this formula we get

$$\Delta^{-1} Q_n(x) \Delta^m g(x) \Big|_{-\infty}^{+\infty} = Q_n(x) \cdot \Delta^{m-1} g(x) \Big|_{-\infty}^{+\infty}$$
$$- \Delta^{-1} \left[\Delta Q_n(x) \cdot \Delta^{m-1} g(x+1) \right]_{-\infty}^{+\infty}$$

The first term on the right hand side vanishes due to condition (c) of the theorem of Charlier. Comparing the term which

remains, i. e.

$$-\Delta^{-1}\left[\Delta Q_n(x)\Delta^{m-1}g(x+1)\right]_{-\infty}^{+\infty} = \begin{cases} 0 & \text{for } n \neq m \\ 1 & \text{for } n = m \end{cases}$$

with the biorthogonality condition

$$\Delta^{-1}\left[Q_{n-1}(x+1)\Delta^{m-1}g(x+1)\right]_{-\infty}^{+\infty} = \begin{cases} 0 & \text{for } n \neq m \\ 1 & \text{for } n = m \end{cases}$$

we conclude that

$$\Delta Q_n(x) = -Q_{n-1}(x+1)$$

This theorem enables us to find the terms of the n th polynomial by taking the negative of the integral of the $(n-1)$ th polynomial, except for the constant of integration. Following the suggestion in our first chapter, we may also determine this constant. We have

$$Q_n(x) = -\Delta^{-1}Q_{n-1}(x+1)\Big|_0^x + C$$

and the simple biorthogonality condition

$$\Delta^{-1}Q_n(x)g(x)\Big|_{-\infty}^{+\infty} = 0.$$

It follows that

$$\Delta^{-1}\left[-\Delta^{-1}Q_{n-1}(x+1)+C\right]_0^x \; g(x)\Big|_{-\infty}^{+\infty} = 0$$

and solving for C we get

$$C = \frac{\Delta^{-1}\left[\Delta^{-1}Q_{n-1}(x+1)\right]_0^x g(x)\Big|_{-\infty}^{+\infty}}{\Delta^{-1}g(x)\Big|_{-\infty}^{+\infty}}.$$

We may therefore determine the polynomials $Q_n(x)$ from the polynomials next preceding by the formula

$$Q_n(x) = -\Delta^{-1}Q_{n-1}(x+1)\Big|_0^x + \frac{\Delta^{-1}\left[\Delta^{-1}Q_{n-1}(x+1)\right]_0^x g(x)\Big|_{-\infty}^{+\infty}}{\Delta^{-1}g(x)\Big|_{-\infty}^{+\infty}}$$

If we adopt the Charlier notation

$$\mathcal{E}_m = \sum_{-\omega}^{+\omega} x^m g(x) = \Delta^{-1} x^m g(x) \Big|_{-\omega}^{+\omega}$$

and the common notation $x^{(m)} = x(x-1)(x-2)\cdots\cdots(x-m+1)$ and observe that $Q_0(x) = 1/\mathcal{E}_0$ and that

$$\Delta^{-1} x^{(m)} = \frac{x^{(m+1)}}{m+1}$$

we may obtain the polynomials $Q_1(x)$, $Q_2(x)$, ... without much computation as follows:

$$Q_1(x) = -\Delta^{-1} Q_0(x+1) \Big|_0^x + \frac{\Delta^{-1}\left[\Delta^{-1} Q_0(x+1)\right]_0^x g(x) \Big|_{-\omega}^{+\omega}}{\Delta^{-1} g(x) \Big|_{-\omega}^{+\omega}}$$

$$= -\frac{x}{\mathcal{E}_0} + \frac{\mathcal{E}_1}{\mathcal{E}_0^2}$$

$$Q_2(x) = -\Delta^{-1} Q_1(x+1) \Big|_0^x + \frac{\Delta^{-1}\left[\Delta^{-1} Q_1(x+1)\right]_0^x g(x) \Big|_{-\omega}^{+\omega}}{\Delta^{-1} g(x) \Big|_{-\omega}^{+\omega}}$$

$$= \frac{(x+1)^{(2)}}{\lfloor 2}\,\mathcal{E}_0 - \frac{\mathcal{E}_1 (x+1)}{\mathcal{E}_0^2} + \frac{2\mathcal{E}_1^2 + \mathcal{E}_1\mathcal{E}_0 - \mathcal{E}_2\mathcal{E}_0}{\lfloor 2\,\mathcal{E}_0^3}$$

$$\text{or } \lfloor 2\,\mathcal{E}_0^3 Q_2(x) = \mathcal{E}_0^2 x^2 - \mathcal{E}_0 x(2\mathcal{E}_1 - \mathcal{E}_0) + 2\mathcal{E}_1^2 - \mathcal{E}_2\mathcal{E}_0 - \mathcal{E}_1\mathcal{E}_0$$

$$Q_3(x) = -\Delta^{-1} Q_2(x+1) \Big|_0^x + \frac{\Delta^{-1}\left[\Delta^{-1} Q_2(x+1)\right]_0^x g(x) \Big|_{-\omega}^{+\omega}}{\Delta^{-1} g(x) \Big|_{-\omega}^{\infty}}$$

$$= -\frac{(x+2)^{(3)}}{\lfloor 3\,\mathcal{E}_0} + \frac{\mathcal{E}_1 (x+2)^{(2)}}{\lfloor 2\,\mathcal{E}_0^2} - \frac{(2\mathcal{E}_1^2 + \mathcal{E}_1\mathcal{E}_0 - \mathcal{E}_2\mathcal{E}_0)(x+2)}{\lfloor 2\,\mathcal{E}_0^3}$$

$$+ \frac{\mathcal{E}_3\mathcal{E}_0^2 - 3\mathcal{E}_2\mathcal{E}_0^2 - 6\mathcal{E}_2\mathcal{E}_1\mathcal{E}_0 + 6\mathcal{E}_1^3 + 6\mathcal{E}_1^2\mathcal{E}_0 + 2\mathcal{E}_1\mathcal{E}_0^2}{\lfloor 3\,\mathcal{E}_0^4}$$

$$\text{or } \lfloor 3\,\mathcal{E}_0^4 Q_3(x) = -\mathcal{E}_0^3 x^3 + 3\mathcal{E}_0^2 x^2 (\mathcal{E}_1 \mathcal{E}_0)$$

$$- \mathcal{E}_0 x(2\mathcal{E}_0^2 - 6\mathcal{E}_1\mathcal{E}_0 - 3\mathcal{E}_2\mathcal{E}_0 + 6\mathcal{E}_1^2) + \mathcal{E}_0^2\mathcal{E}_3 + 3\mathcal{E}_2\mathcal{E}_0^2 + 2\mathcal{E}_1\mathcal{E}_0^2$$

$$- 6\mathcal{E}_2\mathcal{E}_1\mathcal{E}_0 - 6\mathcal{E}_1^2\mathcal{E}_0 + 6\mathcal{E}_1^3$$

. . . . . .

These results differ slightly from those obtained by Charlier in his article. This is due to the definition for differences used by Charlier, viz.:

$$\Delta g(x) = g(x) - g(x-1)$$

while we have used the definition

$$\Delta g(x) = g(x+1) - g(x) .$$

Denoting the difference

$$g(x) - g(x-1) \text{ by } \delta g(x)$$

Charlier determines a set of polynomials $T_n(x)$ satisfying the conditions,

$$\delta^{-1}\left[T_n(x)\,\delta^m g(x)\right] \begin{cases} = 0 & \text{for } m \neq n \\ = 1 & \text{for } m = n \end{cases}$$

As a consequence by paralleling the reasoning above one proves easily that the $T_n(x)$ satisfy the recurrence relation

$$T_n(x+1) - T_n(x) = -T_{n-1}(x) .$$

By using this relation and the fact that

$$\delta^n g(x+n) = \Delta^n g(x)$$

it can be shown without much difficulty that

$$T_n(x+n-1) = Q_n(x)$$

The theorem proved in Ch. 1, par. 2, could no doubt be paralleled by using finite difference theory. Since the method of procedure is obvious there seems to be no need of taking it up in detail.

We have succeeded in showing in this chapter that the problem of determining the constants for the Charlier Type B series closely parallels the work of the first chapter and that these constants are readily obtained by using the biorthogonality conditions for finite differences.

CHAPTER IV

Polynomials Connected with the Pearson Difference Equation

1. In Chapter II we referred to certain solutions $f(x)$ of the Pearson differential equation and noted that graphically, these functions represented types of curves used in statistical work. Paralleling this work, we would expect to find that a difference equation similar in composition to the Pearson differential equation would have as solutions functions $q(x)$ which could be used to represent data consisting of discrete variates.

Carver, in an article in the "Handbook of Mathematical Statistics,"* suggests the use of a difference equation corresponding to the Pearson differential equation, i. e.:

$$\Delta u_x = -\frac{a_0 + a_1 x}{b_0 + b_1 x + b_2 x^2 + b_3 x^3 + \cdots} u_x,$$

a difference equation with a numerator of the first and denominator of any desired degree in x. If we confine our work to a denominator of degree at most of the second in x, we should be able to obtain results comparing very favorably with those obtained in the second chapter.

An illustration of a solution of this difference equation found in Charlier's article "Ueber die Darstellung willkürlicher Funktionen,"[2] is the well known Poisson exponential function

$$\psi(x) = \frac{e^{-\lambda} \lambda^x}{x!}$$

[1]H. C. Carver, "Frequency Curves," Handbook of Mathematical Statistics (H. L. Rietz, Editor), Chapter VII, pp. 111-114.

[2]C. V. L. Charlier, op. cit. p. 33.

This function satisfies the difference equation

$$\Delta u_x = \frac{\lambda - x - 1}{x+1}\, u_x$$

and this equation is recognized as a special form of the Pearson difference equation. If we take the successive differences of this Poisson exponential function, we find that these give rise to a unique set of polynomials. These polynomials may be written in the following form:

$$Q_1(x) = \lambda - (x+1),$$

$$Q_2(x) = \lambda^2 - 2\lambda(x+2) + (x+2)(x+1),$$

or making use of the usual difference notation for $x^{(m)} = x(x-1)(x-2)\cdots\cdots(x-m+1)$, we write

$$Q_2(x) = \lambda^2 - 2\lambda(x+2) + (x+2)^{(2)},$$

$$Q_3(x) = \lambda^3 - 3\lambda^2(x+3) + 3\lambda(x+3)^{(2)} - (x+3)^{(3)},$$

or $$Q_3(x-3) = \lambda^3 - 3\lambda^2 x + 3\lambda x^{(2)} - x^{(3)},$$

. . . . . .

$$Q_n(x) = \lambda^n - {}_nC_1\lambda^{n-1}(x+n) + {}_nC_2\lambda^{n-2}(x+n)^{(2)} + \cdots + (-1)^n(x+n)^{(n)}$$

or $$Q_n(x-n) = \lambda^n - {}_nC_1\lambda^{n-1}x + {}_nC_2\lambda^{n-2}x^{(2)} + \cdots + (-1)^n x^{(n)}.$$

These polynomials have the same form as that for the binomial expansion $(\lambda - x)^n$, particularly if we use the difference notation for representing powers of x. In other words, we might look upon the nth polynomial as being defined as

$$Q_n(x-n) = [\lambda - x^{(1)}]^n$$

A careful examination of these polynomials brings out the fact that consecutive ones are related to each other, viz., that we have,

$$\Delta Q_n(x) = -nQ_{n-1}(x+1) .$$

This relation is similar to the one found for Hermite polynomials.

The fact that the Charlier Type A. series in Chapter II consisted of successive derivatives and that the derivatives of the solutions of the Pearson differential equation led to a system of polynomials definitely related to one another, gave rise to the theory developed in that chapter. We found that it was not necessary in this theory to consider the form of the solution of the equation, but that a set of general polynomials could be set up which satisfied all the properties of the special polynomials. The Charlier Type B series consists of successive differences of a function $g(x)$ and it is quite natural for us to suspect that we can develop for the solutions of the Pearson difference equation a corresponding theory on polynomials.

This question of obtaining a system of polynomials from the solutions of the Pearson difference equation

$$(1) \qquad \Delta u_x = \frac{a_0 + a_1 x}{b_0 + b_1 x + b_2 x^2} u_x .$$

numerator of the first degree and denominator of the second degree, will concern us in this chapter. We shall further show that these polynomials are related to one another by means of first and second order difference relations and by means of recurrence relations involving the $(n+1)$ th, n th and $(n-1)$ th polynomials, and shall illustrate these equations with the Poisson exponential function.

2. For convenience denote the numerator $(a_o + a_1 x)$ in equation (1) by N_x and the denominator $(b_o + b_1 x + b_2 x^2)$ by D_x. We may then define a set of polynomials by the following theorem:

THEOREM: *If* u_x *is a non-identically zero solution of*

$$\Delta u_x = \frac{N_x}{D_x} u_x$$

then $\frac{1}{u_x} D_x D_{x+1} D_{x+2} \cdots D_{x+n-1} \Delta^n u_x$ *is a polynomial of degree at most* n, i. e. $Q_n(x)$.

The proof will proceed by mathematical induction. If we recall the formula for the difference of a product

$$\Delta [u_x v_x] = v_x \Delta u_x + u_{x+1} \Delta v_x = v_x \Delta u_x + (u_x + \Delta u_x) \Delta v_x ,$$

we obtain by differencing

$$D_x \Delta u_x = N_x u_x = Q_1(x) u_x$$

the equation

$$D_{x+1} \Delta^2 u_x + \Delta u_x \Delta D_x = [Q_1(x) + \Delta Q_1(x)] \Delta u_x + u_x \cdot \Delta Q_1(x).$$

Using the value for Δu_x from the original difference equation and multiplying the equation through by D_x, we obtain:

$$D_x D_{x+1} \Delta^2 u_x = [N_x Q_1(x) + D_x \Delta Q_1(x) + N_x \Delta Q_1(x) + N_x \Delta D_x] u_x$$

Since the coefficient of u_x is a polynomial of degree at most 2 in x, we write

$$D_x D_{x+1} \Delta^2 u_x = Q_2(x) u_x .$$

Let us now assume that the statement holds for $m \leq n$, i. e.

$$D_x D_{x+1} D_{x+2} \cdots D_{x+n-1} \Delta^n u_x = Q_n(x) u_x .$$

Differencing both sides of this equation gives us

$$D_x D_{x+1} D_{x+2} \cdots D_{x+n-1} \Delta^{n+1} u_x + (\Delta^n u_x + \Delta^{n+1} u_x)(\Delta D_x D_{x+1} D_{x+2} \cdots D_{x+n-1})$$

$$= Q_n(x) u_x + (u_x + \Delta u_x) \Delta Q_n(x).$$

Now

$$\Delta D_x D_{x+1} \cdots D_{x+n-1} = D_{x+1} D_{x+2} \cdots D_{x+n} - D_x D_{x+1} \cdots D_{x+n-1} =$$

$$= (D_{x+1} \cdots D_{x+n-1})(D_{x+n} - D_x).$$

Hence by the definition of $Q_n(x)$

$$\Delta\left[D_x D_{x+1} \cdots D_{x+n-1} \Delta^n u_x\right] = \frac{D_{x+n} - D_x}{D_x} Q_n(x) u_x .$$

Substituting these values in the above equation as well as the value for Δu_x from (1) and multiplying by D_x, the equation reduces to

$$D_x D_{x+1} D_{x+2} \cdots D_{x+n} \Delta^{n+1} u_x = \left[N_x Q_n(x) + D_x \Delta Q_n(x) + N_x \Delta Q_n(x) - D_{x+n} Q_n(x) + D_x Q_n(x)\right] u_x .$$

The coefficient of u_x on the right hand side is a polynomial of degree at most n in x. We therefore conclude that

$$D_x D_{x+1} \cdots D_{x+n} \Delta^{n+1} u_x = Q_{n+1}(x) u_x .$$

We have also succeded in deriving a relation similar to relation (I) of Chapter II, i. e.

$$Q_{n+1}(x) = (N_x + D_x - D_{x+n}) Q_n(x) + (N_x + D_x) \Delta Q_n(x), \tag{XI}$$

a relation which shows that the $(n+1)$ th polynomial is made up of the n th polynomial and the difference of the n th polynomial. This relation differs from relation (I) in the fact that the coefficient of $\Delta Q_n(x)$ is $N_x + D_x$ instead of D_x. This change seems to be connected with the fact that the original difference equation

$$D_x \Delta u_x = N_x u_x$$

can also be written

$$u_{x+1} = \frac{N_x + D_x}{D_x} u_x .$$

Formula (XI) may also be written

$$Q_{n+1}(x) = (N_x + D_x) Q_n(x+1) - D_{x+n} Q_n(x) \tag{XI'}$$

since $Q_n(x) + \Delta Q_n(x) = Q_{n+1}(x)$.

It seems advisable to adopt a notation for the term

$$D_x D_{x+1} D_{x+2} \cdots D_{x+n-1}$$

since it will continue to be involved in the work that is to follow. The difference notation $x^{(m)} = x(x-1)(x-2) \cdots (x-m+1)$ suggests that we use the symbol $D_x^{(n)}$, i. e.

$$D_x^{(n)} = D_x D_{x-1} D_{x-2} \cdots D_{x-n+1}.$$

Then we will have

$$D_x D_{x+1} D_{x+2} \cdots D_{x+n-1} = D_{x+n-1}^{(n)}$$

and

$$\Delta D_{x+n-1}^{(n)} = D_{x+n} D_{x+n-1} \cdots D_{x+2} D_{x+1} - D_{x+n-1} D_{x+n-2} \cdots D_{x+1} D_x = (D_{x+n} - D_x) \cdot D_{x+n-1}^{(n-1)}.$$

3. We may also define the general polynomials $Q_n(m, x)$ where m is any integer, by means of a theorem as follows:

THEOREM: *If u_x is a non-identically zero solution of the difference equation* (1), *then*

$$\frac{D_{x-m+n-1}^{(n)} \Delta^n \left[D_{x-1}^{(m)} u_x \right]}{D_{x-1}^{(m)} u_x}$$

is a polynomial $Q_n(m, x)$, and $Q_n(m, x)$ is at most of degree n in x. In particular if $m = n$, we have

$$\frac{1}{u_x} \Delta^n \left[D_{x-1}^{(n)} u_x \right]$$

is a polynomial in x of degree at most n.

This theorem may be proved by using the following lemma:

LEMMA: *If u_x satisfy the difference equation* (1), *then $D_{x-1}^{(m)} u_x$, where m is any positive integer, satisfies a difference equation of the same type, viz.:*

$$\Delta \left[D_{x-1}^{(m)} u_x \right] = \frac{D_{x-1}^{(m)} u_x \left[N_x + D_x - D_{x-m} \right]}{r_{x-m}}.$$

The proof proceeds easily by mathematical induction.

For $m = 1$ we have

$$\Delta[D_{x-1}u_x] = D_x \Delta u_x + u_x \Delta D_{x-1}$$

$$= N_x u_x + \Delta D_{x-1} u_x$$

$$= D_{x-1} u_x \left[\frac{N_x + D_x - D_{x-1}}{D_{x-1}}\right],$$

For $m = 2$, we get

$$\Delta[D_{x-2} D_{x-1} u_x] = D_{x-1} \Delta[D_{x-1} u_x] + D_{x-1} u_x \Delta D_{x-2}$$

$$= D_{x-2} D_{x-1} u_x \left[\frac{N_x + \Delta D_{x-1} + \Delta D_{x-2}}{D_{x-2}}\right],$$

or $$\Delta[D_{x-1}^{(2)} u_x] = D_{x-1}^{(2)} u_x \left[\frac{N_x + D_x - D_{x-2}}{D_{x-2}}\right].$$

Let us assume that it holds for the mth case, i. e.

$$\Delta[D_{x-1}^{(m)} u_x] = D_{x-1}^{(m)} u_x \left[\frac{N_x + D_x - D_{x-m}}{D_{x-m}}\right].$$

Then

$$\Delta[D_{x-m-1} D_{x-1}^{(m)} u_x] = D_{x-m} \Delta[D_{x-1}^{(m)} u_x] + D_{x-1}^{(m)} u_x \Delta D_{x-m-1}$$

$$= D_{x-1}^{(m)} u_x [N_x + D_x - D_{x-m} + D_{x-m} - D_{x-m-1}]$$

$$= D_{x-1}^{(m+1)} u_x \left[\frac{N_x + D_x - D_{x-m-1}}{D_{x-m-1}}\right].$$

Making use of this lemma in proving the last theorem, we note that

$$\Delta^2[D_{x-1}^{(m)} u_x] = D_{x-1}^{(m)} u_x \frac{[N_x + D_x - D_{x-m}]}{D_{x-m} D_{x-m+1}}$$

or $$D_{x-m} D_{x-m+1} \Delta^2[D_{x-1}^{(m)} u_x] = D_{x-1}^{(m)} u_x [N_x + D_x - D_{x-m}]$$

and in general that

$$\Delta^n\left[D_{x-1}^{(m)} u_x\right] = \frac{D_{x-1}^{(m)} u_x \left[Q_n(m,x)\right]}{D_{x-m+n-1}^{(n)}}$$

or $$D_{x-m+n-1}^{(n)} \Delta^n\left[D_{x-1}^{(m)} u_x\right] = D_{x-1}^{(m)} u_x \; Q_n(m,x).$$

In particular, if $m = n$, we define the polynomials $Q_n(n,x)$ as $\Delta^n\left[D_{x-1}^{(n)} u_x\right] = Q_n(n,x)\, u_x$ which relation is of interest because the Δ^n has no D_x as multiplier. Any result derived for the polynomials $Q_n(x) = \frac{1}{u_x} D_{x+n-1}^{(n)} \Delta^n u_x$ where u_x is a solution of the difference equation (1) can now be extended to the polynomials $Q_n(m,x) = \frac{\Delta^n\left[D_{x-1}^{(m)} u_x\right]}{D_{x-1}^{(m-n)} u_x}$ by replacing N_x by $(N_x + D_x - D_{x-m})$ and D_x by D_{x-m}. For example, relation (XI) becomes

$$Q_{n+1}(m+1,x) = (N_x + D_x - D_{x-m+n-1})\, Q_n(m+1,x) + (N_x + D_x)\, \Delta Q_n(m+1,x) \qquad (XI_m)$$

and when $m = n$, this relation reduces to

$$Q_{n+1}(n+1,x) = (N_x + D_x - D_{x-1}) Q_n(n+1,x) + (N_x + D_x)\Delta Q_n(n+1,x) \qquad (XI_n)$$

$$= (N_x + \Delta D_{x-1}) Q_n(n+1,x) + (N_x + D_x)\Delta Q_n(n+1,x).$$

4. In analogy with the work of chapter II, we next proceed to find a recurrence relation involving the $(n+1)$ th, n th and $(n-1)$ th of the polynomials $Q(x)$. We take the n th difference of both sides of the equation

$$D_x \Delta u_x = N_x u_x$$

by making use of the formula for the n th difference of a product

$$\Delta^n [u_x v_x] = v_x \Delta^n u_x + n \Delta v_x \Delta^{n-1} u_{x+1} + \frac{n(n-1)}{2!} \Delta^2 v_x \Delta^{n-2} u_{x+2} + \cdots .$$

We then obtain the equation

$$D_x \Delta^{n+1} u_x + n \Delta D_x \Delta^n u_{x+1} + \frac{n(n-1)}{2!} \Delta^2 D_x \Delta^{n-1} u_{x+2} =$$

$$N_x \Delta^n u_x + n \Delta N_x \Delta^{n-1} u_{x+1} .$$

$\Delta^3 D_x$ and $\Delta^2 N_x$ being equal to zero. Multiplying through by $D_{x+n}^{(n)}$ we get

$$D_{x+n}^{(n)} \Delta^{n+1} u_x + n D_{x+n}^{(n)} \Delta D_x \Delta^n u_{x+1} + \frac{n(n-1)}{2!} D_{x+n}^{(n)} \Delta^2 D_x \Delta^{n-1} u_{x+2}$$

$$= N_x D_{x+n}^{(n)} \Delta^n u_x + n D_{x+n}^{(n)} \Delta N_x \Delta^{n-1} u_{x+1}$$

But $u_{x+1} = u_x + \Delta u_x$ and $u_{x+2} = u_x + 2\Delta u_x + \Delta^2 u_x$.

Substituting these values in the last equation and using the defiinition for the polynomials $Q_n(x)$, we obtain:

$$Q_{n+1}(x) u_x + \frac{n \Delta D_x}{D_x} \left[D_{x+n} Q_n(x) + Q_{n+1}(x) \right] u_x$$

$$+ \frac{n(n-1)}{2!} \frac{D_{x+1}}{D_{x+1}} \frac{\Delta^2 D_x}{D_x} \left[D_{x+n}^{(2)} Q_{n-1}^{(x)} + 2 D_{x+n} Q_n(x) + Q_{n+1}(x) \right] u_x$$

$$= \frac{N_x}{D_x} D_{x+n} Q_n(x) u_x + \frac{n D_{x+n} \Delta N_x}{D_x} \left[D_{x+n-1} Q_{n-1}(x) + Q_n(x) \right] u_x$$

Dividing through by u_x and collecting like terms, this expression reduces to

$$\left[1+\frac{n\Delta D_x}{D_x}+\frac{n(n-1)}{2}\frac{\Delta^2 D_x}{D_x}\right]Q_{n+1}(x)+\left[\frac{nD_{x+n}\Delta D_x}{D_x}+\frac{n(n-1)\Delta^2 D_x D_{x+n}}{D_x}\right.$$

$$\left.-\frac{N_x D_{x+n}}{D_x}-\frac{nD_{x+n}\Delta N_x}{D_x}\right]Q_n(x)+\left[\frac{n(n-1)}{2!D_x}D_{x+n-1}D_{x+n}\Delta^2 D_x\right.$$

$$\left.-\frac{nD_{x+n-1}D_{x+n}\Delta N_x}{D_x}\right]Q_{n-1}(x)=0.$$

Now we know that

$$u_{x+n}=u_x+n\Delta u_x+\frac{n(n-1)}{2!}\Delta^2 u_x+\cdots\cdots$$

and so we may write D_{x+n} and N_{x+n} in this same form, i. e.

$$D_{x+n}=D_x+n\Delta D_x+\frac{n(n-1)}{2!}\Delta^2 D_x,$$

$$\Delta D_{x+n}=\Delta D_x+n\Delta^2 D_x,$$

and $$N_{x+n}=N_x+n\Delta N_x$$

the third and higher differences of D_x and the second and higher differences of N_x being equal to zero. The coefficient of $Q_{n+1}(x)$ reduces to $\frac{2D_{x+n}}{D_x}$ and the coefficient of $Q_n(x)$ also reduces to a simpler form. Dividing through by $\frac{2D_{x+n}}{D_x}$ we finally get the recurrence relation:

$$Q_{n+1}(x)+(n\Delta D_{x+n-1}-N_{x+n})Q_n(x)$$

(XII)

$$+nD_{x+n-1}\left[\frac{(n-1)}{2}\Delta^2 D_x-\Delta N_x\right]Q_{n-1}(x)=0$$

i. e. the $(n+1)$ th polynomial may be obtained from the n th and $(n-1)$ th polynomials.

In Chapter II we found that relations (I) and (II) were identical for the first two terms, and as a consequence we equated the third terms and obtained a relation between the derivative of

a polynomial $P_n(x)$ and the polynomial preceding it. In order that we may obtain a similar expression for the difference polynomials, we must change the appearance of formula (XII).

By lowering the degree in formula (XI) from n to $n-1$ and solving for $D_{x+n-1}\, Q_{n-1}(x)$ we find that

$$D_{x+n-1}\, Q_{n-1}(x) = (N_x + D_x)\, Q_{n-1}(x+1) - Q_n(x).$$

Substitution of this relation in formula XII gives

$$Q_{n+1}(x) = (N_{x+n} - n\Delta D_{x+n-1})\, Q_n(x)$$

$$+ n\left[\Delta N_x - \frac{(n-1)}{2}\Delta^2 D_x\right]\left[(N_x + D_x)\, Q_{n-1}(x+1) - Q_n(x)\right]$$

or $$Q_{n+1}(x) = \left[N_{x+n} - n\Delta N_x - n\Delta D_{x+n-1} + \frac{n(n-1)}{2}\Delta^2 D_x\right] Q_n(x)$$

$$+ n\left[\Delta N_x - \frac{(n-1)}{2}\Delta^2 D_x\right](N_x + D_x)\, Q_{n-1}(x+1).$$

Just as in Chapter IV, paragraph 3, the coefficient of $Q_n(x)$ reduces and becomes the same as the coefficient of $Q_n(x)$ in formula (XI) and we have

$$Q_{n+1}(x) = (N_x + D_x - D_{x+n})\, Q_n(x)$$

(XII′)

$$+ n\left[\Delta N_x - \frac{(n-1)}{2}\Delta^2 D_x\right](N_x + D_x)\, Q_{n-1}(x+1).$$

We therefore conclude that

(XIII) $$\Delta Q_n(x) = n\left[\Delta N_x - \frac{(n-1)}{2}\Delta^2 D_x\right] Q_{n-1}(x+1),$$

a relation expressing the difference of a polynomial $Q_n(x)$ in terms of the next preceding polynomial in $(x+1)$, i. e.

$Q_{n-1}(x+1)$. For the polynomial $Q_n(n,x)$, formula (XII) may be written in the form

$$(\mathrm{XIII}_n)\quad \Delta Q_n(n,x) = n\left[\Delta N_x + \frac{(n+1)}{2}\Delta^2 D_x\right] Q_{n-1}(n,x+1),$$

this relation being obtained by replacing N_x by $(N_x + D_x - D_{x-n})$ and D_x by D_{x-n}.

Formula XIII which was just derived is the general form of the relation we found to hold for the Poisson exponential function polynomials, i. e.

$$\Delta Q_n(x) = -nQ_{n-1}(x+1).$$

We find further that these polynomials satisfy a special form of (XI), i. e.

$$Q_{n+1}(x) + (x+n+1-\lambda)Q_n(x) - \lambda\Delta Q_n(x) = 0$$

and for formula (XII) we get the special form

$$Q_{n+1}(x) + (x+2n+1-\lambda)Q_n(x) + n(x+n)Q_{n-1}(x) = 0$$

This recurrence relation is also similar to the one given for Laguerre polynomials.

5. Turning now to the problem of obtaining a second order difference relation for the polynomials $Q_n(x)$, we proceed to difference formula (XI), i. e.

$$Q_{n+1}(x) = (N_x + D_x - D_{x+n})Q_n(x) + (N_x + D_x)\Delta Q_n(x)$$

and get

$$\Delta Q_{n+1}(x) = (\Delta N_x + \Delta D_x - \Delta D_{x+n}) Q_n(x)$$

$$+ (N_{x+1} + D_{x+1} - D_{x+n+1}) \Delta Q_n(x)$$

$$+ (\Delta N_x + \Delta D_x) \Delta Q_n(x) + (N_{x+1} + D_{x+1}) \Delta^2 Q_n(x).$$

Substituting for $\Delta Q_{n+1}(x)$ the value

$$(n+1)\left[\Delta N_x - \frac{n}{2}\Delta^2 D_x\right]\left[Q_n(x) + \Delta Q_n(x)\right]$$

found in formula (XIII), gives us

$$(n+1)\left[\Delta N_x - \frac{n}{2}\Delta^2 D_x\right]\left[Q_n(x) + \Delta Q_n(x)\right] =$$

$$\left[\Delta N_x + \Delta D_x - \Delta D_{x+n}\right] Q_n(x) + \left[N_{x+1} + D_{x+1} - D_{x+n+1}\right] \Delta Q_n(x)$$

$$+ \left[\Delta N_x + \Delta D_x\right] \Delta Q_n(x) + \left[N_{x+1} + D_{x+1}\right] \Delta^2 Q_n(x).$$

Collecting the coefficients of like terms and simplifying them, we finally get

$$(N_{x+1} + D_{x+1}) \Delta^2 Q_n(x) + \left[N_{x-n+1} - (n-1)\Delta D_x\right] \Delta Q_n(x)$$

(XIV)

$$- n\left[\Delta N_x - \frac{(n-1)}{2}\Delta^2 D_x\right] Q_n(x) = 0,$$

a relation very similar in form to formula (IV) and consisting of the first and second differences of the polynomial $Q_n(x)$. This relation when applied to the Poisson exponential function gives

$$\lambda \Delta^2 Q_n(x) + (\lambda - x - 1)\Delta Q_n(x) + n Q_n(x) = 0,$$

an equation which can be checked by substituting the value of the general Poisson polynomial in it.

The extension of formula (XIV) to the polynomials $Q_n(m, x)$ and $Q_n(n, x)$ by making the proper substitutions for N_x

and D_x results in the following expressions:

$$(N_{x+1}+D_{x+1})\Delta^2 Q_n(m,x)$$

$$(\mathrm{XIV}_m)\quad +\left[N_{x-n+1}+D_{x-n+1}-D_{x-m-n+1}-(n-1)\Delta D_{x-m}\right]\Delta Q_n(m,x)$$

$$-n\left[\Delta N_x+\Delta D_x-\Delta D_{x-m}-\left(\frac{n-1}{2}\right)\Delta^2 D_{x-m}\right]Q_n(m,x)=0,$$

which may also be written as:

$$(N_{x+1}+D_{x+1})\Delta^2 Q_n(m,x)$$

$$+\left[N_{x-n+1}+(m-n+1)\Delta D_x-\frac{m(m+1)}{2}\Delta^2 D_x\right]\Delta Q_n(m,x)$$

$$-n\left[\Delta N_x-\frac{n-2m-1}{2}\Delta^2 D_x\right]Q_n(m,x)=0$$

In particular if $m=n$ we have:

$$(N_{x+1}+D_{x+1})\Delta^2 Q_n(n,x)+$$

$$(\mathrm{XIV}_n)\quad +\left[N_{x-n+1}+\Delta D_x-\frac{n(n+1)}{2}\Delta^2 D_x\right]\Delta Q_n(n,x)$$

$$-n\left[\Delta N_x+\frac{(n+1)}{2}\Delta^2 D_x\right]Q_n(n,x)=0.$$

6. The next set of relations we shall derive are recurrence relations for the polynomials $Q_n(m,x)$ and $Q_n(n,x)$. In the lemma proved in this chapter we found that

$$\Delta\left(D_{x-1}^{(m+1)}u_x\right)=\left[D_{x-1}^{(m)}u_x\right]\left[N_x+D_x-D_{x-m-1}\right].$$

Taking the n th difference of both sides of the equation gives:

$$\Delta^{n+1}\left(D_{x-1}^{(m+1)}u_x\right)=(N_x+D_x-D_{x-m-1})\Delta^n\left[D_{x-1}^{(m)}u_x\right]$$

$$+n(\Delta N_x-\Delta D_x-\Delta D_{x-m-1})\Delta^{n-1}\left[D_x^{(m)}u_{x+1}\right],$$

the second difference of the trinomial $(N_x+D_x-D_{x-m-1})$ being equal to zero. Multiplying this last expression through by

$D_{x-m+n-1}^{(n+1)}$ and substituting for $D_{x-m+n-1}^{(n+1)} \Delta^{n+1} D_{x-1}^{(m+1)} u_x$ the value $D_{x-1}^{(m+1)} Q_{n+1}(m,x) u_x$, we get

$$D_{x-1}^{(m+1)} Q_{n+1}(m+1,x) u_x = (N_x + D_x - D_{x-m-1}) D_{x-1}^{(m+1)} Q_n(m,x) u_x$$
$$+ n(\Delta N_x + \Delta D_x - \Delta D_{x-m-1}) D_x^{(m+2)} \left(\frac{N_x + D_x}{D_x}\right) Q_{n-1}(m,x+1) u_x.$$

Dividing through by $D_{x-1}^{(m+1)} u_x$ we get a recurrence relation involving the polynomials $Q_{n+1}(m+1,x)$, $Q_n(m,x)$ and $Q_{n-1}(m,x+1)$, i. e.

$$Q_{n+1}(m+1,x) = (N_x + D_x - D_{x-m-1}) Q_n(m,x)$$

(XV_m)

$$+ n\left[\Delta N_x + (m+1)\Delta^2 D_x\right](N_x + D_x) Q_{n-1}(m,x+1).$$

For $m = n$, this expression reduces to:

$$Q_{n+1}(n+1,x) = (N_x + D_x - D_{x-n-1}) Q_n(n,x)$$

(XV_n)

$$+ n\left[\Delta N_x + (n+1)\Delta^2 D_x\right](N_x + D_x) Q_{n-1}(n,x+1).$$

7. Another form of this relation is obtained by substituting the value found in $(XIII_n)$ for $Q_{n-1}(n,x+1)$, i. e.

$$Q_{n-1}(n,x+1) = \frac{1}{n\left[\Delta N_x + \frac{n+1}{2}\Delta^2 D_x\right]} \Delta Q_n(n,x),$$

in formula (XV_n), which gives

$$Q_{n+1}(n+1,x) = (N_x + D_x - D_{x-n-1}) Q_n(n,x)$$

(XVI)

$$+ (N_x + D_x) \frac{\Delta N_x + (n+1)\Delta^2 D_x}{\Delta N_x + \frac{(n+1)}{2}\Delta^2 D_x} \Delta Q_n(n,x),$$

a relation very similar to formula (VI).

8. There remains one more formula in Chapter II for which we have not yet found a parallel in this chapter, i. e. formula VII. To obtain this parallel expression, we difference formula (XVI), thereby obtaining:

$$\Delta Q_{n+1}(n+1,x)=(\Delta N_x+\Delta D_x-\Delta D_{x-n-1})\,Q_n(n,x)$$

$$+(N_{x+1}+D_{x+1}-D_{x-n})\,\Delta Q_n(n,x)$$

$$+\frac{\Delta N_x+(n+1)\Delta^2 D_x}{\Delta N_x+\frac{(n+1)}{2}\Delta^2 D_x}\left[(\Delta N_x+\Delta D_x)\Delta Q_n(n,x)+(N_{x+1}+D_{x+1})\Delta^2 Q_n(n,x)\right].$$

In formula (XIV_n) we found a value for

$$(N_{x+1}+D_{x+1})\Delta^2 Q_n(n,x)$$

which when substituted in this last expression gives us:

$$\Delta Q_{n+1}(n+1,x)=(\Delta N_x+\Delta D_x-\Delta D_{x-n+1})Q_n(n,x)+(N_{x+1}+D_{x+1}-D_{x-n})\Delta Q_n(n,x)$$

$$+\frac{\Delta N_x+(n+1)\Delta^2 D_x}{\Delta N_x+\frac{(n+1)}{2}\Delta^2 D_x}\left[\Delta N_x+\Delta D_x+\frac{n(n+1)}{2}\Delta^2 D_x-N_{x-n+1}-\Delta D_x\right]\Delta Q_n(n,x)$$

$$+n\frac{\Delta N_x+(n+1)\Delta^2 D_x}{\Delta N_x+\frac{(n+1)}{2}\Delta^2 D_x}\left[\Delta N_x+\frac{(n+1)}{2}\Delta^2 D_x\right]Q_n(n,x).$$

Collecting coefficients we get

$$\Delta Q_{n+1}(n+1,x)=\left[\Delta N_x+n\Delta N_x+\Delta D_x-\Delta D_{x-n-1}+n(n+1)\Delta^2 D_x\right]Q_n(n,x)$$

$$+\left[N_{x+1}+D_{x+1}-D_{x-n}\right]Q_n(n,x)$$

$$+\frac{\Delta N_x+(n+1)\Delta^2 D_x}{\Delta N_x+\frac{(n+1)}{2}\Delta^2 D_x}\left[\Delta N_x-N_{x-n+1}+\frac{n(n+1)}{2}\Delta^2 D_x\right]\Delta Q_n(n,x)$$

and by simplifying the coefficients this expression finally reduces to the formula

$$\Delta Q_{n+1}(n+1,x)=(n+1)\left[\Delta N_x+(n+1)\Delta^2 D_x\right] Q_n(n,x)$$

$$+\left\{N_{x+1}+(n+1)\Delta D_x-\frac{n(n+1)}{2}\Delta^2 D_x\right.$$

(XVII)

$$\left.-\left[\frac{\Delta N_x+(n+1)\Delta^2 D_x}{\Delta N_x+\left(\frac{n+1}{2}\right)\Delta^2 D_x}\right]\left[N_{x-n}-\frac{n(n+1)}{2}\Delta^2 D_x\right]\right\}\Delta Q_n(n,x)$$

a relation which is also similar in form to formula VII.

Before concluding this chapter, we might examine the character of the polynomials $Q_n(n,x)$ when the original function is the Poisson exponential function $\psi(x)=\frac{e^{-\lambda}\lambda^x}{x!}$.

We find these polynomials to have the following form:

$$\frac{1}{\psi(x)}\Delta\frac{xe^{-\lambda}\lambda^x}{x!}=Q_1(1,x)=\lambda-x,$$

$$\frac{1}{\psi(x)}\Delta^2\frac{x^{(2)}e^{-\lambda}\lambda^x}{x!}=Q_2(2,x)=\lambda^2-2\lambda x+x^{(2)}$$

$$\frac{1}{\psi(x)}\Delta^3\frac{x^{(3)}e^{-\lambda}\lambda^x}{x!}=Q_3(3,x)=\lambda^3-3\lambda^2 x+3\lambda x^{(2)}-x^{(3)},$$

- -

$$\frac{1}{\psi(x)}\Delta^n\frac{x^{(n)}e^{-\lambda}\lambda^x}{x!}=Q_n(n,x)=\lambda^n-n\lambda^{n-1}x+\frac{n(n-1)}{2!}\lambda^{n-2}x^{(2)}+\cdots+(-1)^n x^{(n)},$$

$$=\left[\lambda-x^{(1)}\right]^n=Q_n(x-n).$$

Substituting the proper values for N_x and D_x in formula (XIV_n) we get

$$\lambda\Delta^2 Q_n(n,x)+(\lambda-x+n-1)\Delta Q_n(n,x)+n\,Q_n(n,x)=0$$

In the same way we find for formula (XI_n) the relation

$$Q_{n+1}(n+1,x)=(\lambda-x+n)Q_n(n,x)+\lambda\Delta Q_n(n,x)$$

and for formula (XVII), the reduced relation

$$\Delta Q_{n+1}(n+1,x)=-(n+1)Q_n(n,x),$$

which is somewhat like the relation obtained for $(XIII_n)$.
We might call attention to the fact that these polynomials are identical with the polynomials obtained by Charlier[1] satisfying the relations

$$\delta^{-1}\left[T_n(x)\,\delta^m\,\frac{e^{-\lambda}\,\lambda^x}{x!}\right]_{-\infty}^{+\infty} = 0 \text{ for } m\neq n$$
$$= 1 \text{ for } m=n$$

9. Summarizing the results of this chapter, we have found that if the general solution $g(x)$ of the difference equation

$$\Delta u_x=\frac{a_0+a_1x}{b_0+b_1x+b_2x^2}\,u_x$$

is used as the generating function $g(x)$ in the Charlier Type B series, that the successive differences give rise to two general types of polynomials which we defined as follows:

$$Q_n(x)=\frac{1}{u_x}\,D_x^{(n)}\Delta^{n+1}u_x$$

and

$$Q_n(n,x)=\frac{1}{u_x}\,\Delta^n D_{x-n}^{(n)}\,u_x\,.$$

With the aid of the properties of the Δ operator, we derived a set of relations and equations for these polynomials of the following form:

---

[1] C. V. L. Charlier: "Ueber die Darstellung willkürlicher Funktionen," p. 34.

(XI) $$Q_{n+1}(x)=(N_x+D_x-D_{x+n})Q_n(x)+(N_x+D_x)\Delta Q_n(x),$$

(XI$_n$) $$Q_{n+1}(n+1,x)=(N_x+\Delta D_{x-1})Q_n(n+1,x)+(N_x+D_x)\Delta Q_n(n,x),$$

(XII) $$Q_{n+1}(x)=(N_{x+n}-n\Delta D_{x+n-1})Q_n(x) + nD_{x+n-1}\left[\Delta N_x-\frac{(n-1)}{2}\Delta^2 D_x\right]Q_{n-1}(x),$$

(XII$'$) $$Q_{n+1}(x)=(N_x+D_x-D_{x+n})Q_n(x) + n\left[\Delta N_x-\frac{(n-1)}{2}\Delta^2 D_x\right](N_x+D_x)Q_{n-1}(x+1),$$

(XIII) $$\Delta Q_n(x)= n\left[\Delta N_x-\frac{(n-1)}{2}\Delta^2 D_x\right]Q_{n-1}(x+1),$$

(XIII$_n$) $$\Delta Q_n(n,x)= n\left[\Delta N_x+\frac{n+1}{2}\Delta^2 D_x\right]Q_{n-1}(n,x+1),$$

(XIV) $$(N_{x+1}+D_{x+1})\Delta^2 Q_n(x)+\left[N_{x-n+1}-(n-1)\Delta D_x\right]\Delta Q_n(x) - n\left[\Delta N_x-\frac{(n-1)}{2}\Delta^2 D_x\right]Q_n(x)=0,$$

(XIV$_n$) $$(N_{x+1}+D_{x+1})\Delta^2 Q_n(n,x)+\left[N_{x-n+1}+\Delta D_x-\frac{n(n+1)}{2}\Delta^2 D_x\right]\Delta Q_n(n,x) - n\left[\Delta N_x+\frac{n+1}{2}\Delta^2 D_x\right]Q_n(n,x)=0,$$

(XV$_n$) $$Q_{n+1}(n+1,x)=(N_x+D_x-D_{x-n-1})Q_n(n,x) + n\left[\Delta N_x+(n+1)\Delta^2 D_x\right](N_x+D_x)Q_{n-1}(n,x+1),$$

(XVI) $$Q_{n+1}(n+1,x)=(N_x+D_x-D_{x-n-1})Q_n(n,x) +(N_x+D_x)\frac{\Delta N_x+(n+1)\Delta^2 D_x)}{\Delta N_x+\frac{(n+1)}{2}\Delta^2 D_x}\Delta Q_n(n,x),$$

$$(\text{XVII})\quad \Delta Q_{n+1}(n+1,x)=(n+1)\left[\Delta N_x+(n+1)\Delta^2 D_x\right]Q_n(n,x)$$
$$+\left\{N_{x+1}+(n+1)\Delta D_x-\frac{n(n+1)}{2}\Delta^2 D_x\right.$$
$$\left.-\left[\frac{\Delta N_x+(n+1)\Delta^2 D_x}{\Delta N_x+\frac{(n+1)}{2}\Delta^2 D_x}\right]\left[N_{x\cdot n}-\frac{n(n+1)}{2}\Delta^2 D_x\right]\right\}\Delta Q_n(n,x).$$

Each of these formulas corresponds and is similar to a formula found in Chapter II. In fact, it seems probable that if we developed the formulas in this present chapter from the equation

$$\frac{\Delta u_x}{\Delta x}=\frac{N_x}{D_x}u_x$$

and permitted the Δ_x to approach zero as a limit, the formulas of Chapter II would result, the above formulas being the case where $\Delta_x = 1$.

E. H. Hildebrandt

A NEW FORMULA FOR PREDICTING THE SHRINKAGE OF THE COEFFICIENT OF MULTIPLE CORRELATION

By

DR. R. J. WHERRY
Cumberland University, Lebanon, Tennessee

With the perfection of the Doolittle Method for the solution of the constant values necessary for the multiple correlation and prediction technique, we may expect a constant increase in the use of this method in statistical practice. Theoretical statisticians have recognized for some time however that the multiple correlation coefficient, derived from a large number of independent variables, is apt to be deceptively large due to chance factors. When prediction equations derived in this manner are applied to subsequent sets of data, there is apt to be a rather large shrinkage in the resulting correlation coefficient obtained, as compared with the original observed multiple correlation coefficient. In order to avoid over optimism it is necessary to have some equation which will predict the most probable value of this shrinkage. The development of such a formula is the purpose of this paper.

The most promising formula of this type so far developed is the B. B. Smith formula, presented by M. J. B. Ezekial at the December, 1928, meeting of the American Mathematical Society held at Chicago. This formula is

$$\bar{R}^2 = 1 - \frac{1-R^2}{1-\frac{M}{N}} = \frac{NR^2-M}{N-M} \tag{1}$$

where $\bar{R}$ = the estimated correlation obtaining in the universe
R = the observed multiple correlation coefficient
M = the number of *independent* variables
N = the number of observations (the statistical population).

This formula was evidently developed by B. B. Smith by an application of the method of least squares as follows (the derivation is that of the author, since he could not find it given elsewhere):

The customary formula for the coefficient of multiple correlation may be written in the form

$$R^2 = 1 - \frac{s_o^2}{\sigma_o^2} \tag{2}$$

where

$$s_o^2 = \frac{\Sigma v^2}{N} \tag{3}$$

where

$$v = x_o - \bar{x}_o \tag{4}$$

The method of least squares, however, says that the most probable value of the standard error of estimate is not that given in equation (3) but

$$\bar{s}_o^2 = \frac{\Sigma v^2}{N-M} = \frac{N}{N-M} \cdot s_o^2 \tag{5}[1]$$

Now, if we substitute the value of (5) in place of (3) in equation (2), we have at once

[1]See Merriman, Method of Least Squares, John Wiley & Sons, London, 8th Edition, pp. 80-82. Also see derivation later in this paper.

$$(6) \qquad \bar{R}^2 = 1 - \frac{s_o^2}{\sigma_o^2} \cdot \frac{N}{N-M}$$

and since, by (2) above, we have $\frac{s_o^2}{\sigma_o^2}$ equal to $(1 - R^2)$, we have

$$(7) \qquad \bar{R}^2 = 1 - \frac{N(1-R^2)}{N-M} = 1 - \frac{1-R^2}{1-\frac{M}{N}}$$

which is, exactly, the B. B. Smith formula (1).

This formula has been widely used during the last few years, but up until recently had not been subjected to much critical examination. However, in a recent article in the *Journal of Educational Psychology*[1], S. C. Larson actually tested the formula empirically on some data obtained from the Mississippi Survey conducted by M. V. O'Shea, obtaining the results indicated in the tables and graphs below, and on the basis of which he reached the following conclusion:

"The Smith Shrinkage-Reduction formula parallels all of the empirical findings but quite consistently gives values which are in excess of those obtained under present experimental conditions." This meant that the Smith formula predicted shrinkages consistently greater than those actually obtained.

It was in view of this reported empirical difference that the writer started his attempt to derive the Smith formula and hit on the method given above. The question at once arose in the writer's mind as to why, when the standard error of estimate had been corrected to correspond to the most probable value by a least squares criterion, the standard deviation of the dependent variable had not been treated in the same fashion.

1"The Shrinkage of the Coefficient of Multiple Correlation," Jan., 1931, pp. 45-55.

Merriman, whose formula we used above in correcting the standard error of estimate (5), likewise, and by identical reasoning, shows that the most probable value of the standard deviation of the dependent variable existing in the universe, should really be represented by the following relationships:

Where

$$\sigma_o^2 = \frac{\Sigma x_o^2}{N} \tag{8}$$

we find

$$\bar{\sigma}_o^2 = \frac{\Sigma x_o^2}{N-1} = \sigma_o^2 \cdot \frac{N}{N-1} \tag{9}$$

which reduces formula (6) to the form

$$\bar{R}^2 = 1 - \frac{s_o^2}{\sigma_o^2} \cdot \frac{\frac{N}{N-M}}{\frac{N}{N-1}} \tag{10a}$$

and when the same substitution is made as in step (7) above, we have

$$\bar{R}^2 = \frac{(N-1)R^2-(M-1)}{N-M} \tag{10b}$$

which is, by a more correctly applied criterion of least squares, the formula we have been seeking, and is a closer approximation than that given by the Smith formula.

The reasons for the substitutions made above in our formulae may not be entirely clear to all readers, so we now present the derivations of the formulae given in (5) and (9) above. The derivations given here are directly adapted from those of Merriman referred to above, but have been translated into the customary statistical notation whenever possible.

First, let us consider the derivation of the value in (9). As

stated in (8) the most customary form of Sigma is

$$\sigma_o^2 = \frac{\Sigma x_o^2}{N} \tag{8}$$

where

$$x_o = x - M_x. \tag{11}$$

Each value x_o has a certain error, however, due to the fact that the value of the mean is merely the most probable value, not the true value. So for each x_o value there is a small unknown error δx_o, so that if we take $\bar{x}_o$ to be the true value of a deviation we have

$$\bar{x}_o = x_o + \delta x_o \tag{12}$$

and, squaring and summating, disregarding the terms involving second power delta terms as small in comparison with the first power terms, we have

$$\Sigma \bar{x}_o^2 = \Sigma x_o^2 + 2\Sigma x_o \delta x_o \tag{13}$$

Now, by the laws of probability, we know that the probability of the occurrence of an error $\bar{x}_o$, whose measure of precision is "h," is

$$\Pi = h d\bar{x}\, \pi^{-\frac{1}{2}} e^{-\bar{x}^2 h^2} \tag{14}$$

multiplying both sides of this equation by $\bar{x}^2$ and summating between the limits plus and minus infinity, we have

$$\Sigma \Pi \bar{x}^2 = \int_{-\infty}^{+\infty} h \bar{x}^2 \pi^{-\frac{1}{2}} e^{-h^2 \bar{x}^2} d\bar{x} = \frac{1}{2h^2} \tag{15}$$

and since $\Sigma \Pi \bar{x}^2$ is the same as $\frac{\Sigma \bar{x}^2}{N}$, since in our work we assume the weight of each value $\bar{x}$, for each of the N observations, to be $\frac{1}{N}$, we have

$$\frac{\Sigma \bar{x}^2}{N} = \frac{1}{2h^2} \tag{16}$$

or

$$\Sigma \bar{x}^2 = \frac{N}{2h^2} \tag{16a}$$

Likewise, if we let

$$2\Sigma x_o \delta x_o = u^2 \tag{17}$$

the probability of the system of errors, u^2, is

$$\Pi' = h\, du\, \pi^{-\frac{1}{2}} e^{-u^2h^2} \tag{18}$$

and the mean of all of the possible values of u^2 is

$$\frac{h}{\pi^{\frac{1}{2}}} \int_{-\infty}^{+\infty} u^2 e^{-h^2u^2} du = \frac{1}{2h^2} \tag{19}$$

and this must be taken as the best attainable value of u^2. But it was shown that the quantity $\frac{1}{2h^2}$ is equal to $\frac{\Sigma \bar{x}^2}{N}$ (16). Hence

$$\Sigma \bar{x}^2 = \Sigma x^2 + \frac{\Sigma \bar{x}^2}{N} \tag{20}$$

from which

$$\bar{\sigma}_o^2 = \frac{\Sigma \bar{x}^2}{N} = \frac{\Sigma x^2}{N-1} = \sigma_o^2 \cdot \frac{N}{N-1} \tag{9}$$

which was to be proved.

To derive (5) we proceed in much the same manner. After our normal equations have been solved for the most probable values of $\beta_{o1}, \beta_{o2}, \beta_{o3}, \cdots \beta_{om}$ for our set of data, we know that these are not the true values, but that they err by small unknown corrections $\delta\beta_{o1}, \delta\beta_{o2}, \delta\beta_{o3}, \cdots \delta\beta_{om}$, the corresponding true values for the universe being $(\beta_{o1} + \delta\beta_{o1})$, $(\beta_{o2} + \delta\beta_{o2})$, $(\beta_{o3} + \delta\beta_{o3})$, . . . $(\beta_{om} + \delta\beta_{om})$.

Now, if we substitute the most probable values of the Betas in our original observation equations, they will not reduce to zero, bus will leave small residuals $v_1, v_2, \ldots v_N$, thus

$$\bar{x}_{o_1} - x_{o_1} = \beta_{o1} x_{1_1} + \beta_{o2} x_{2_1} + \beta_{o3} x_{3_1} + \cdots + \beta_{om} x_{m_1} - x_{o_1} = v_1$$

$$\bar{x}_{o_2} - x_{o_2} = \beta_{o1} x_{1_2} + \beta_{o2} x_{2_2} + \beta_{o3} x_{3_2} + \cdots + \beta_{om} x_{m_2} - x_{o_2} = v_2$$

. . . . . .

$$\bar{x}_{o_M} - x_{o_M} = \beta_{o1} x_{1_N} + \beta_{o2} x_{2_N} + \beta_{o3} x_{3_N} + \cdots + \beta_{om} x_{m_N} - x_{o_N} = v_N$$

while if the corresponding true values be substituted, we obtain

$$(\beta_{o1} + \delta\beta_{o1}) x_{1_1} + (\beta_{o2} + \delta\beta_{o2}) x_{2_1} + \cdots + (\beta_{om} + \delta\beta_{om}) x_{m_1} - x_{o_1} = \bar{v}_1$$

$$(\beta_{o1} + \delta\beta_{o1}) x_{1_2} + (\beta_{o2} + \delta\beta_{o2}) x_{2_2} + \cdots + (\beta_{om} + \delta\beta_{om}) x_{m_2} - x_{o_2} = \bar{v}_2$$

. . . . . .

$$(\beta_{o1} + \delta\beta_{o1}) x_{1_N} + (\beta_{o2} + \delta\beta_{o2}) x_{2_N} + \cdots + (\beta_{om} + \delta\beta_{om}) x_{m_N} - x_{o_N} = \bar{v}_N$$

Subtracting each of the former equations from the latter, we obtain

$$v_1 + \delta\beta_{01} x_{1_1} + \delta\beta_{02} x_{2_1} + \cdots + \delta\beta_{0m} x_{m_1} = \bar{v}_1$$

$$v_2 + \delta\beta_{01} x_{1_2} + \delta\beta_{02} x_{2_2} + \cdots + \delta\beta_{0m} x_{m_2} = \bar{v}_2$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$v_N + \delta\beta_{01} x_{1_N} + \delta\beta_{02} x_{2_N} + \cdots + \delta\beta_{0m} x_{m_N} = \bar{v}_N$$

Now the principle of least squares provides that $\Sigma \bar{v}^2$ shall be made a minimum to give the most probable values of β_{01}, β_{02}, . . . β_{0m}, and by the solution of the normal equations by the Doolittle method its minimum value is found to be Σv^2. From the residual equations we may find the relationship existing between the values $\Sigma \bar{v}^2$ and Σv^2. Thus, if we square each equation immediately above and then summate we have (if we neglect squares and products of the delta values as small in comparison with the first powers):

$$(21) \quad \Sigma v^2 + 2\delta\beta_{01}\Sigma x_1 v + 2\delta\beta_{02}\Sigma x_2 v + \cdots + 2\delta\beta_{0m}\Sigma x_m v = \Sigma \bar{v}^2$$

which we may write as

$$(22) \quad \Sigma v^2 + u_1^2 + u_2^2 + \cdots + u_m^2 = \Sigma \bar{v}^2$$

Now, by analogous reasoning to that in steps (14), (15), and

(16), we may set

$$\Sigma \bar{v}^2 = \frac{N}{2h^2} \tag{23}$$

Further, if there be but one independent variable, there will be but one $2\delta\beta_{ox}\Sigma x_x v$ and its value by the same process used in steps (18) and (19) can be shown to be

$$u_x^2 = \frac{1}{2h^2} \tag{24}$$

and since that is true whichever unknown quantity be considered, the values of each u_x^2 value must be $\frac{1}{2h^2}$; and as there are M of these values the above equation (22) becomes

$$\Sigma v^2 + \frac{M}{2h^2} = \frac{N}{2h^2}$$

from which

$$h = \sqrt{\frac{N-M}{2\Sigma v^2}} \tag{25}$$

Therefore, from the constant relationship which exists between the value "h" and the Probable Error, we have

$$PE_{\bar{v}} = 0.6745\sqrt{\frac{\Sigma v^2}{N-M}} \tag{26}$$

and therefore, by the relationship existing between the probable error and the standard deviation we have at once

$$\sigma_{\bar{v}}^2 = \bar{s}_o^2 = \frac{\Sigma v^2}{N-M} \tag{5}$$

which was to be proved.

The next step was to test out the formula empirically. This was done by using Larson's material, with the results indicated in the tables below, and in the graphs which show the same

set of facts, but which make the results much more apparent.

An inspection of the tables and graphs will show at once that the new formula predicts what will actually happen much more accurately than the Smith formula did. In graph 1, for example, the agreement is so good that the results appear almost to have been a regression line fit to the particular set of data.

It was to have been expected that if the formula actually predicted the most probable values of the correlations obtaining in the universe that the errors incurred by the use of the formula would be normally distributed around zero as a mean value. Graph 3 presents a comparison of the error curves obtained by use of the Smith and the Wherry prediction formulae, together with an approximation to the normal curve. As a further and more scientific check the criteria for a normal curve as set forth by Rietz[1] were applied to the data. His criteria are

$$\mu_1 = 0, \beta_1 = \frac{\mu_3^2}{\mu_2^3} = 0, \quad \beta_2 = \frac{\mu_4}{\mu_2^2} = 3; \qquad \text{where } \mu_n = \frac{\Sigma x^n}{N}$$

The results for the two formulae are given below.

(Results based on an expectancy of zero)

	Smith Formula	Wherry Formula
μ_1	.00138	.00038
β_1	.223	.025
β_2	3.004	3.703

[1]Rietz, H. L. Mathematical Statistics, Carus Mathematical Monograph No. 3, Mathematical Association of America, Chicago 1927, pp. 58-59.

It is apparent therefore that the Wherry formula gave much better results for both the first criterion (mean error) and the second criterion (skewness), but that the excess was greater for the Wherry formula than for the Smith formula. However, one cannot quarrel too much with getting errors actually smaller than would be expected by assuming normality. Even this superiority is seen to be fictitious if the distributions are measured from their own means rather than from an expected mean of zero. When this is done, which is the manner in which the criteria are customarily used, we have

(Results based on means of distributions)

	Smith Formula	Wherry Formula
α_1	.000	.000
β_1	1.712	.025
β_2	5.524	3.753

Thus, we find that the Smith distribution has, in reality, even a greater excess than does the Wherry formula, but has it at a point farther removed from the desired value.

SUMMARY AND CONCLUSIONS

1. Larson has shown that the theoretically expected shrinkage is an empirical fact.

2. Larson has shown that the Smith formula, when tested empirically, consistently over-estimates this shrinkage as determined empirically.

3. It has been demonstrated that the new Wherry formula,

both by a least squares criterion and by actual application, is more nearly true than the corresponding Smith formula.

4. The correct formula for the shrunken coefficient of multiple correlation is

$$\bar{R}^2 = \frac{(N-1)R^2 - (M-1)}{N-M}$$

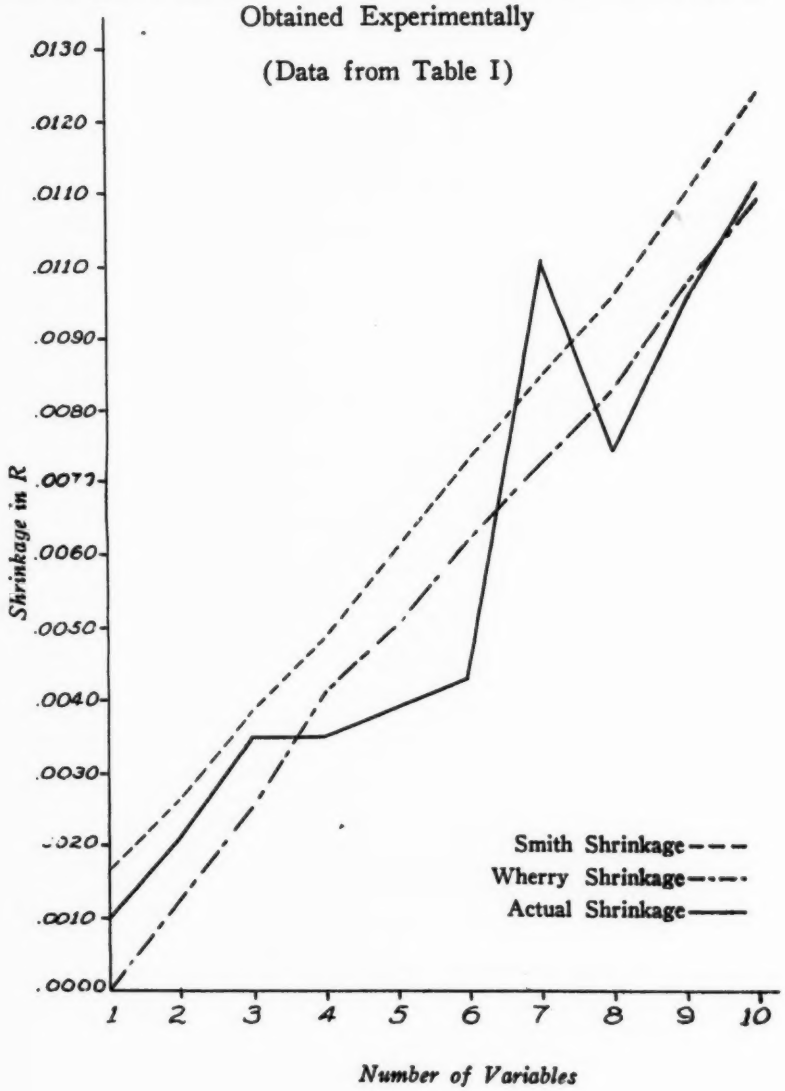
where $\bar{R}$ = the estimated correlation obtaining in the universe
R = the observed coefficient of multiple correlation
M = the number of *independent* variables
and N = the number of observations (statistical population).

R. J. Wherry

GRAPH 1.

Shrinkage as Obtained by Use of the Formulae and Also as Obtained Experimentally

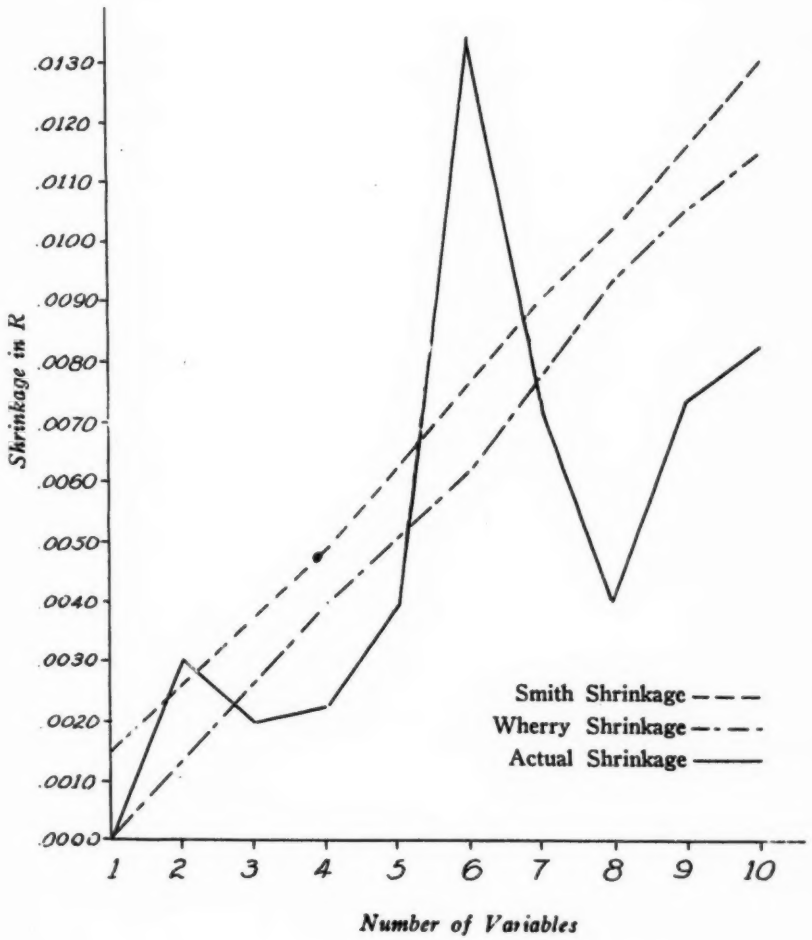
(Data from Table I)



GRAPH 2

Shrinkage as Obtained by Use of the Formulae and Also as Obtained Experimentally

(Data from Table II)



GRAPH 3

Ogive Showing the Distribution of Error in Predicting Shrinkage

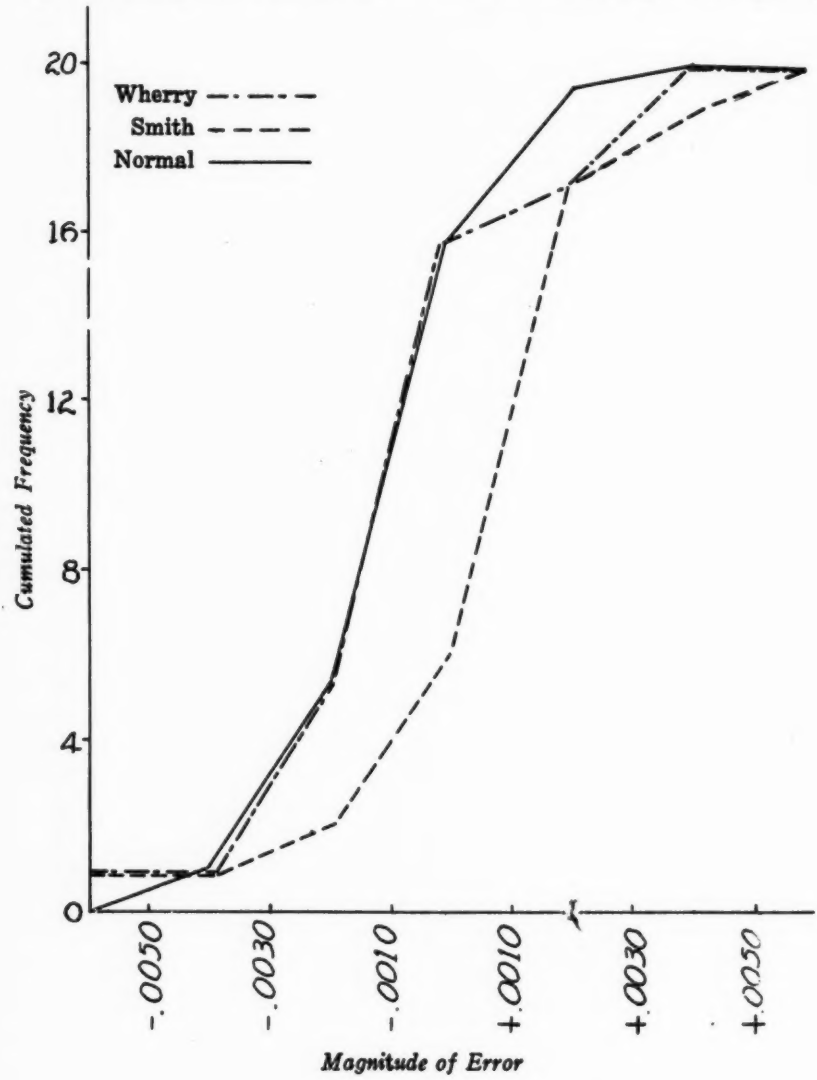


TABLE I*

Showing the Actual Shrinkage in $\mathcal{R}$ Found When the Prediction Equation Found on One Group of Subjects Is Applied to a Comparable Group, Together with the Shrinkage of $\mathcal{R}$ as Indicated by the Smith and Wherry Formulae. The Statistical Population (n) Is 200 Throughout.

m	1	2	3	4	5	6	7	8	9	10
$\mathcal{R}$	.7042	.7794	.7834	.7872	.7880	.7907	.7929	.7941	.7944	.7945
Actual Shrinkage	.0000	.0021	.0036	.0036	.0060	.0044	.0102	.0075	.0097	.0113
Shrinkage by Smith formula	.0017	.0026	.0038	.0049	.0062	.0074	.0085	.0097	.0110	.0123
Shrinkage by Wherry formula	.0000	.0013	.0025	.0040	.0049	.0062	.0073	.0085	.0098	.0111

*The article by Larson reported the values for the Smith formula erroneously, due to a misconception of the meaning of m. Those in the present tables are the correct values.

TABLE II

Showing for a Second Set of Groups the Same Facts as Obtain in Table I

m	1	2	3	4	5	6	7	8	9	10
R	.7402	.7759	.7813	.7826	.7847	.7858	.7859	.7863	.7868	.7869
Actual Shrinkage	.0000	.0031	.0019	.0023	.0041	.0133	.0073	.0042	.0074	.0083
Shrinkage by Smith formula	.0015	.0026	.0038	.0051	.0063	.0076	.0089	.0102	.0115	.0129
Shrinkage by Wherry formula	.0000	.0013	.0026	.0038	.0051	.0063	.0076	.0089	.0105	.0115

TABLE III

Showing the Mean Error Attained by the Use of the Smith and Wherry Shrinkage Formulae.

Formula	Table I	Table II	Tables I and II
Smith	.00097	.00180	.00138
Wherry	.00018	.00057	.00038
N	10	10	20

THE USE OF THE RELATIVE RESIDUAL IN THE APPLICATION OF THE METHOD OF LEAST SQUARES

By

WALTER A. HENDRICKS
Junior Biologist, Bureau of Animal Industry,
U. S. Department of Agriculture.

The method of least squares offers a precise method of fitting a curve describing the relation between two or more related, measurable variables, but certain criteria must be fulfilled to justify its application. First, the type of equation selected for fitting must be the true mathematical expression of the law governing the relationship of the variables. Secondly, all errors of measurement, made in obtaining the observed values of the variables when the data were collected, must be distributed according to the well-known laws of probability.[1]

This paper is concerned with the latter of these two criteria. The fundamental theory upon which the method of least squares is based can be found in any text-book on the subject and need not be elaborated upon here. However, it may be well to point out a very pertinent, if somewhat elementary, aspect of the theory which facilitates the ready visualization of the fundamental concepts involved.

---

[1]Steinmetz, C. P. Engineering Mathematics. McGraw-Hill Book Co., New York (1917).

The application of the method of least squares to curve fitting, as ordinarily described in works on the subject, is perfectly analogous to the calculation of the arithmetic mean of a number of measurements made upon a single, constant quantity. This may be easily demonstrated as follows:

Let $Y = f(X)$ describe the relation existing between an independent variable, X, and a dependent variable, Y. If it is desired to find the most probable value of the dependent variable when X has some definite value, X_1, the most direct method of procedure would be to make a number of measurements of Y at this value of X and calculate their arithmetic mean, provided, of course, that the errors of measurement were distributed according to the laws of probability in a normal frequency distribution. According to the elementary theory of statistics, the most probable value of the dependent variable, $f(X_1)$, would be such that the sum of the squares of the deviations of the actual measurements from this value would be a minimum.

If X is conceived to be varying in value so rapidly that it is impossible to make more than one measurement of Y at any value of X, this direct method can not be employed. However, the most probable value of $f(X_1)$ can still be determined. Let $Y_1, Y_2, Y_3, \ldots Y_n$ each represent a measured value of Y at values $X_1, X_2, X_3, \ldots X_n$, respectively, of the independent variable. Since the errors of measurement are assumed to be distributed according to the law of chance, an error of a given magnitude is equally likely to occur at any value of X. In other words, exactly the same errors would be made in obtaining one measurement of each of the quantities $Y_1, Y_2, Y_3, \ldots Y_n$ as if $f(X_1)$ were measured n times. These errors may, therefore, be considered as having been made in measuring a single, constant quantity. Therefore, if $f(X)$ denotes the most probable value of Y at any value of X and Y denotes the corresponding observed value, the most probable values of the dependent variable which can be calculated from any set of

data are such that the sum of the squares of the differences, $f(X)-Y$, is a minimum.

It is important to bear in mind that this conception of the distribution of errors of measurement is justified only when an error of a given magnitude is equally likely to occur at any value of X . In actual practice it often happens that this ideal condition is not realized. The magnitude of the errors of measurement is often influenced by the magnitude of the quantity which is being measured. In obtaining the live weights of animals at different ages, for example, it is common practice to use a less delicate balance in making the weighings as the animals become larger, and the magnitude of the errors of measurement increases as the sensitivity of the balance decreases. Other factors which tend to increase the magnitude of the errors may also be in operation. The error, or rather the unreliability, of the weight of a 1,000 pound steer would be greater than that of a 100 pound calf, even though an equally sensitive balance were used in making both weighings, because of a greater content of material in the digestive tract and excretory organs and the increased effect of the movements of the animal.

It is highly probable that in many fields of investigation such disturbing influences are encountered more frequently than the ideal conditions which justify the application of the method of least squares as ordinarily described.

Pearl and Reed recognized the need for modifying the application of the method of least squares to compensate for changes in the probability of the occurrence of an error of any given magnitude and suggested, as stated by Pearl,[1] that it would be more logical in many instances to employ residuals of the type $\frac{f(X)-Y}{Y}$. The use of such residuals was based on the assumption that if the errors of measurement were expressed as percentages of the

[1]Pearl, Raymond. Studies in Human Biology. Williams & Wilkins, Baltimore (1924).

magnitude of the quantities measured, the percentage errors would be distributed at random according to the law of probability. In many practical problems this assumption appears to be justifiable.

The study herein reported was made to determine the extent of the error made when the method of least squares as ordinarily described is applied to data in which the percentage, rather than the absolute errors of measurement are distributed according to the law of chance.

The writer desired a hypothetical set of errors of measurement which, when expressed as percentages of the quantities measured, would come as near as possible to forming a normal frequency distribution.

TABLE I

Ideal Frequency Distribution of 41 Throws of 12 Dice in Which a Throw of 4, 5, or 6 Points Is Considered a Success.

SUCCESSES	FREQUENCY
2	1
3	2
4	5
5	8
6	9
7	8
8	5
9	2
10	1
	—
Total	41

Mills[1] gives the results of fitting a normal frequency curve to Weldon's distribution of 4096 throws of 12 dice, described by Yule,[2] in which a throw of 4, 5, or 6 points was considered a success. If each frequency, calculated from the fitted curve, is divided by 100 and the results rounded off to whole numbers, the frequency distribution given in Table I is obtained.

If hypothetical errors of measurement are substituted for

TABLE II

Ideal Frequency Distribution of 41 Hypothetical Percentage Errors of Measurement.

ERROR (Per cent of quantity measured)	FREQUENCY
+ 8	1
+ 6	2
+ 4	5
+ 2	8
0	9
− 2	8
− 4	5
− 6	2
− 8	1
Total	41

[1]Mills, F. C. Statistical Methods Applied to Economics and Business. Henry Holt & Co., New York (1924).

[2]Yule, G. Udny. Introduction to the Theory of Statistics. Charles Griffin & Co., Ltd., London (1927).

successes in this frequency table, the resulting distribution may be considered to represent a distribution of random errors of measurement which might be made in obtaining a series of 41 measurements of a variable. The most probable error should obviously be zero. If the total range in magnitude of the errors is assumed to be from + 8 per cent to −8 per cent and the precision of measurement is such that each error differs from the next larger or smaller error by 2 per cent, the distribution of these hypothetical errors of measurement should be as given in Table II.

From the simple equation, $Y = 100\ X^2$, 41 values of Y were calculated, using values of X from 1 to 41, inclusive. Each calculated value of Y was then changed by algebraically subtracting the hypothetical errors of measurement given in Table II. All the percentage errors of each magnitude were arbitrarily distributed as uniformly as possible throughout the data. These altered values of Y will hereafter be termed the "observed" values and the original values, from which they were calculated, the "true" values. The observed values of Y, together with the true values and the assumed errors of measurement from which they were calculated, are given in Table III.

In order to be certain that the errors were actually distributed in such a manner that the probability of the occurrence of a percentage error of any given magnitude was the same at all values of X, the writer employed Pearson's method of square contingency as described by Yule.[1] A 16-cell contingency table was constructed in which the percentage errors were classified according to the values of X at which they occurred. The chi-square test for contingency was applied to this table.

Table IV shows the actual distribution of the percentage errors, together with the corresponding theoretical frequencies. Since there are 4 rows and 4 columns of cells in the table, the number of algebraically independent differences between theoret-

[1]Loc. cit.

TABLE III

Calculation of the Observed Values of Y from the True Values.

X	$100X^2$	ERROR		Y	X	$100X^2$	ERROR		Y
		Per cent	Actual Units				Per cent	Actual Units	
1	100	− 2	− 2	102	22	48400	+ 8	+ 3872	44528
2	400	+ 4	+ 16	384	23	52900	0	0	52900
3	900	0	0	900	24	57600	− 4	− 2304	59904
4	1600	− 4	− 64	1664	25	62500	− 2	− 1250	63750
5	2500	+ 6	+ 150	2350	26	67600	0	0	67600
6	3600	− 6	− 216	3816	27	72900	− 2	− 1458	74358
7	4900	+ 2	+ 98	4802	28	78400	+ 6	+ 4704	73696
8	6400	0	0	6400	29	84100	+ 2	+ 1682	82418
9	8100	− 2	− 162	8262	30	90000	+ 4	+ 3600	86400
10	10000	+ 2	+ 200	9800	31	96100	− 4	− 3844	99944
11	12100	+ 4	+ 484	11616	32	102400	+ 2	+ 2048	100352
12	14400	− 4	− 576	14976	33	108900	0	0	108900
13	16900	− 8	− 1352	18252	34	115600	+ 2	+ 2312	113288
14	19600	+ 2	+ 392	19208	35	122500	− 2	− 2450	124950
15	22500	− 2	− 450	22950	36	129600	0	0	129600
16	25600	0	0	25600	37	136900	+ 4	+ 5476	131424
17	28900	+ 2	+ 578	28322	38	144400	− 6	− 8664	153064
18	32400	+ 4	+ 1296	31104	39	152100	− 2	− 3042	155142
19	36100	− 2	− 722	36822	40	160000	0	0	160000
20	40000	0	0	40000					
21	44100	+ 2	+ 882	42318	41	168100	− 4	− 6724	174824

ical and observed frequencies is $(4-1)(4-1)+1$ or 10. The value of X^2, calculated from the data in Table IV, is 1.3171. The corresponding value of P, which is the probability that as bad, or worse, an agreement between observed and theoretical frequencies could occur from the fluctuations of random sampling is, according to Pearson's Tables,[1] 0.996911 or almost certainty. The percentage errors were, therefore, distributed in such a manner as to be uncorrelated with the values of X at which they were used.

The equation, $Y=AX^2$, was fitted to the hypothetical set of data in Table III by the method of least squares as ordinarily described. If AX^2 represents a calculated value of the dependent variable and Y represents the corresponding observed value, the difference between these two values is AX^2-Y and the square of the difference is $A^2X^4-2AX^2Y+Y^2$. The sum of the squares of all the differences is $A^2\Sigma X^4-2A\Sigma X^2Y+\Sigma Y^2$. The value of this expression will be a minimum when its derivative with respect to A is equal to zero. Differentiating and equating to zero yields the following equations for the determination of A:

$$(1) \qquad 2A\Sigma X^4-2\Sigma X^2Y=0$$

$$(2) \qquad A=\frac{\Sigma X^2Y}{\Sigma X^4}$$

The value of A calculated from the data in Table III by means of equation (2) is 100.6250.

If residuals of the type suggested by Pearl and Reed are employed, A is calculated as follows. Let AX^2 represent a calculated value of the dependent variable, as before, and let Y represent the corresponding observed value. Then the difference between the two values, expressed as a fraction of the observed

---

[1]Pearson, Karl. Tables for Statisticians and Biometricians. Cambridge University Press, London (1924).

TABLE IV

Chi-square test for contingency applied to the distribution of the percentage errors of measurement. The theoretical frequencies for each compartment are given in parentheses.

Value of X	Magnitude of Error (Per cent)				
	0.0 to ±1.9	± 2.0 to ± 3.9	± 4.0 to ± 5.9	± 6.0 and over	Total
1 to 10	2 (2.1951)	4 (3.9024)	2 (2.4390)	2 (1.4634)	10
11 to 20	2 (2.1951)	4 (3.9024)	3 (2.4390)	1 (1.4634)	10
21 to 30	2 (2.1951)	4 (3.9024)	2 (2.4390)	2 (1.4634)	10
31 to 41	3 (2.4146)	4 (4.2926)	3 (2.6829)	1 (1.6097)	11
Total	9	16	10	6	41

$X^2 = 1.3171$

$n' = 10$

$P = 0.996911$

value, is $\frac{AX^2-Y}{Y}$ or $\frac{AX^2}{Y} - 1$. The square of this relative deviation is $\frac{A^2X^4}{Y^2} - \frac{2AX^2}{Y} + 1$ and the sum of the squares of the 41 relative deviations is $A^2\Sigma\frac{X^4}{Y^2} - 2A\Sigma\frac{X^2}{Y} + 41$. This expression will likewise have its minimum value when its derivative with respect to A is equal to zero. Differentiating and equating to zero, as before, leads to the following equations for the determination of A :

$$2A\Sigma\frac{X^4}{Y^2} - 2\Sigma\frac{X^2}{Y} = 0 \tag{3}$$

$$A = \frac{\Sigma\frac{X^2}{Y}}{\Sigma\frac{X^4}{Y^2}} \tag{4}$$

Applying equation (4) to the given set of data gives a value of 99.7573 for A. This value of A is closer to the true value 100, than the value which was calculated by means of equation (2) but the improvement was not as great as might be expected.

It occurred to the writer that if the deviations of the calculated, from the observed, values of the dependent variable were expressed as fractions of the calculated values, a more accurate value of A could be obtained.

The relative deviation expressed in this manner is $\frac{AX^2-Y}{AX^2}$ or $1 - \frac{A^{-1}Y}{X^2}$. The square of this deviation is $1 - \frac{2A^{-1}Y}{X^2} + \frac{A^{-2}Y^2}{X^4}$ and the sum of the squares of the 41 relative deviations is $41 - 2A^{-1}\Sigma\frac{Y}{X^2} + A^{-2}\Sigma\frac{Y^2}{X^4}$. Differentiating this expression with respect to A and equating to zero yields the following equations for the determination of A :

$$2A^{-2}\Sigma\frac{Y}{X^2} - 2A^{-3}\Sigma\frac{Y^2}{X^4} = 0 \tag{5}$$

$$A = \frac{\Sigma\frac{Y^2}{X^4}}{\Sigma\frac{Y}{X^2}} \tag{6}$$

The value of A, calculated from the data by means of equation (6), is 100.1210 which is nearer to the true value than either of the values calculated by the two preceding methods. However, it is evident that equation (6) failed to give results as precise as one would expect, in view of the method by which the observed values of Y were obtained.

The reason for this discrepancy can be made most apparent by returning to the analogy existing between the application of the method of least squares to curve fitting and the calculation of the arithmetic mean of a number of measurements of a single, constant quantity.

Let $m_1, m_2, m_3, \ldots m_n$ represent measured values of the same constant quantity and let their arithmetic mean be represented by M. If each measurement is divided by the arithmetic mean of all the measurements, the resulting distribution of these relative values will be normal if the original measurements were distributed normally. The arithmetic mean of these relative values will obviously be unity.

Let $\frac{m_1}{M}, \frac{m_2}{M}, \frac{m_3}{M}, \ldots \frac{m_n}{M}$ represent the relative values of the measurements. The arithmetic mean of these values is unity. Therefore, the deviation of any relative value, $\frac{m}{M}$, from the mean is $1 - \frac{m}{M}$.

Let it be assumed that the value of the arithmetic mean of the original measurement, M, is unknown and is represented by Z. Then any measurement, m, expressed as a fraction of Z, is $\frac{m}{Z}$. According to the discussion in the two preceding paragraphs, it might appear that Z must have such a value that the sum of the squares of the deviations, $1 - \frac{m}{Z}$, is a minimum. However, this is not the case. It may be demonstrated that the value of the expression $\Sigma\left(1 - \frac{2m}{Z} + \frac{m^2}{Z^2}\right)$, is a minimum when Z has some other value than the arithmetic mean of the original measurements. The sum of the squares of the residuals may be written, $n - 2Z^{-1}\Sigma m + Z^{-2}\Sigma m^2$. Differentiating this expression with respect to Z and equating to zero yields the following

equations for the determination of Z :

$$(7) \qquad 2Z^{-2}\Sigma m - 2Z^{-3}\Sigma m^2 = 0$$

$$(8) \qquad Z = \frac{\Sigma m^2}{\Sigma m}$$

The value of Z , calculated by means of equation (8), is obviously not the arithmetic mean of the original measurements. The fallacy in the deduction of this equation is readily apparent.

Instead of using residuals of the type, $1 - \frac{m}{Z}$, and differentiating the sum of the squares of the residuals with respect to Z , one should use residuals of the type, $V - \frac{m}{Z}$, in which V represents the arithmetic mean of the relative values, $\frac{m}{Z}$, of the measurements. The sum of the squares of the residuals should be differentiated with respect to V . The square of the residual, $V - \frac{m}{Z}$, is $V^2 - \frac{2m}{Z} + \frac{m^2}{Z^2}$ and the sum of the squares of all the residuals may be written $nV^2 - \frac{2V}{Z}\Sigma m + \frac{1}{Z^2}\Sigma m^2$ Differentiating with respect to V and equating to zero yields the following equations for the determination of V :

$$(9) \qquad 2nV - \frac{2}{Z}\Sigma m = 0$$

$$(10) \qquad V = \frac{\frac{1}{Z}\Sigma m}{n}$$

Since the value of V is known to be unity, equation (10) may be written:

$$(11) \qquad n = \frac{1}{Z}\Sigma m$$

from which Z may be readily calculated as follows:

$$(12) \qquad Z = \frac{\Sigma m}{n}$$

Equation (12) is obviously nothing more than the simple formula for the calculation of the arithmetic mean of the original measurements, which is sufficient evidence that the reasoning involved in its deduction is sound.

It is now readily apparent why equation (6) did not yield results which were consistent with the data in Table III. The ratio, $\frac{Y}{AX^2}$, is analogous to the ratio, $\frac{m}{Z}$, and residuals of the type, $V - \frac{Y}{AX^2}$, should have been used in fitting the equation instead of residuals of the type, $1 - \frac{Y}{AX^2}$.[1] The square of the residual, $V - \frac{Y}{AX^2}$, is $V^2 - \frac{2VY}{AX^2} + \frac{Y^2}{A^2X^4}$. The sum of the squares of the 41 residuals is $41V^2 - \frac{2V}{A}\Sigma\frac{Y}{X^2} + \frac{1}{A^2}\Sigma\frac{Y^2}{X^4}$. Differentiating this expression with respect to V and equating to zero yields the following equations for the determination of V:

$$(13) \qquad 82V - \frac{2}{A}\Sigma\frac{Y}{X^2} = 0$$

$$(14) \qquad V = \frac{\frac{1}{A}\Sigma\frac{Y}{X^2}}{41}$$

Substituting the known value, unity, for V in equation (14) yields the following equations for the determination of A.

$$(15) \qquad 41 = \frac{1}{A}\Sigma\frac{Y}{X^2}$$

$$(16) \qquad A = \frac{\Sigma\frac{Y}{X^2}}{41}$$

[1] Residuals of the type, $\frac{AX^2}{Y} - 1$, are analogous to those of the type, $\frac{Z}{m} - 1$, which also lead to incorrect results.

Applying equation (16) to the data in Table III gives A a value of 100.0000, which coincides exactly with the true value from which the data were originally calculated. Equation (16) was, therefore, the correct equation to use in interpreting the data given in Table III. Although the use of residuals of the types, $\frac{AX^2}{Y} - 1$ and $1 - \frac{Y}{AX^2}$, gave better approximations to the true values of A than the use of the simple residuals, $AX^2 - Y$, neither of the two gave results which were entirely in accord with the derivation of the data.

Yule[1] suggested that the geometric mean might often prove useful in comparing the frequency distributions of different sets of data, in which the dispersion of the individual measures about their means was influenced by the magnitude of the means. It appeared to the writer that the use of residuals of the type, $\log AX^2 - \log Y$, might give a good approximation to the true value of A in fitting the given equation. It is evident that the ratio, $\frac{AX^2}{Y}$, approaches unity as the residual, $\log AX^2 - \log Y$, approaches zero.

This logarithmic residual may be written, $\log A + 2 \log x - \log Y$, and its square is $(\log A)^2 + 4(\log X)^2 + (\log Y)^2 + 4(\log A)(\log X) - 2(\log A)(\log Y) - 4(\log X)(\log Y)$. The sum of the squares of the 41 residuals is $41(\log A)^2 + 4\Sigma(\log X)^2 + \Sigma(\log Y)^2 + 4(\log A)\Sigma(\log X) - 2(\log A)\Sigma(\log Y) - 4\Sigma(\overline{\log X} \cdot \overline{\log Y})$. Differentiating this expression with respect to $\log A$ and equating to zero yields the following equations for the determination of A :

$$82(\log A) + 4\Sigma(\log X) - 2\Sigma(\log Y) = 0 \tag{17}$$

$$\log A = \frac{(\log Y) - 2\Sigma(\log X)}{41} \tag{18}$$

[1]Loc. cit.

The value of $\log A$, calculated from the given set of data by means of equation (18), is 1.9997369, which gives A a value of 99.9394. This value of A comes closer to the true value than those calculated by means of residuals of the types, $1-\frac{Y}{AX^2}$ and $\frac{AX^2}{Y}-1$. However, since the use of the geometric mean is not rigorously justified when the distribution of the measures about the arithmetic mean is symmetrical, the use of logarithmic residuals in curve fitting can not give precise results when the errors of measurement are distributed as they were in the given set of data.

In any application of the method of least squares to a practical problem, the procedure of the investigators should be governed by the nature of the data to which it is being applied. In many instances the correct procedure can be deduced by a careful consideration and evaluation of the accuracy of the methods of measurement used in obtaining the data. Unfortunately, however, some sources of error are not always readily apparent at the time the data are collected, and occasionally can not be quantitatively estimated even though they are known to exist. If the nature of the mathematical relationship existing between the dependent and independent variables is known, all that remains is to find the most probable values of the constants in the equation.

A statistical study of the deviations of the observed values of the dependent variable from the corresponding calculated values, obtained after fitting the equation by several different methods, may be of much help in deciding which method of fitting was most consistent with the nature of the data. For example, Table V gives the results of applying the chi-square test for contingency to the distribution of the deviations of the observed values of Y from the calculated values obtained when residuals of the type, AX^2-Y, were used in fitting the equation, $Y=AX^2$, to the data in Table III. The value of P is only 0.005061 and a mere inspection of the table itself shows that large deviations tend to occur more frequently, and small deviations less frequently, as

TABLE V

Chi-square test for contingency applied to the distribution of the deviations of the type, AX^2-Y. The theoretical frequencies for each compartment are given in parentheses.

Value of X	Magnitude of Deviation				
	10 to ± 1999	± 2000 to ± 3999	± 4000 to ± 5999	± 6000 and over	Total
1 to 10	10 (7.3171)	0 (1.2197)	0 (0.9756)	0 (0.4878)	10
11 to 20	10 (7.3171)	0 (1.2197)	0 (0.9756)	0 (0.4878)	10
21 to 30	6 (7.3171)	1 (1.2197)	3 (0.9756)	0 (0.4878)	10
31 to 41	4 (8.0488)	4 (1.3415)	1 (1.0732)	2 (0.5366)	11
Total	30	5	4	2	41

$X^2 = 23.5989$

$n' = 10$

$P = 0.005061$

the values of X increase. If the true nature of the values of Y in Table III were not known in advance, this distribution of the deviations would be sufficient evidence that the method of fitting the equation was not consistent with the accuracy of the measurements made when the data were collected.

Tables VI, VII, and VIII give, respectively, the distributions of the deviations of the types, $\frac{AX^2}{Y} - 1$, $1 - \frac{Y}{AX^2}$, and $\log AX^2 - \log Y$, when the corresponding residuals were used in fitting the equation.[1] The value of D is high in each case, indicating that, although the use of residuals of these types did not give results which were precisely accurate, nevertheless, they yielded values of A which were well within the limits of the probable error to be expected in any practical investigation.

As a matter of fact, this is a rather fortunate circumstance, since the only method of fitting the equation given above which yielded exactly the correct value of A cannot be applied to fitting an equation containing more than one undetermined constant. The applicability of residuals of the types, $1 - \frac{Y}{f(X)}$ and $\log f(X) - \log Y$ is also somewhat limited. However, any equation which can be fitted by the method of least squares at all can still be fitted when residuals of the type, $\frac{f(X)}{Y} - 1$, are employed.

SUMMARY AND CONCLUSIONS

The method of least squares can be a more valuable tool in statistical work when the fundamental theory upon which the method is based is taken into consideration. The use of residuals of the type, $f(X) - Y$, is probably justified in fewer practical

---

[1]The distribution of the deviations obtained when the equation was fitted to the data by means of equation (16) is identical with the distribution of the errors given in Table IV.

problems than the use of residuals of some other form. The type of residual to be employed should be governed by the nature of the data to which the method of least squares is being applied.

The use of relative residuals of the type suggested by Pearl and Reed may be of much value in many instances but will not give results which are precisely accurate, even though the distribution of the percentage errors of measurement is strictly normal. The results can be improved by expressing the deviations of the observed from the calculated values of the dependent variable as fractions of the calculated, rather than the observed, value.[1]

The use of logarithmic residuals may give more accurate results than the use of residuals of the type suggested by Pearl and Reed, even though the distribution of the percentage errors of measurement is normal.

The chi-square test for contingency may be of much help in selecting the type of residual most consistent with the errors of measurement made in obtaining the data when sufficient information regarding the accuracy of the measurements is not available.

---

[1]Residuals of this type have been used by Hendricks, Lee, and Titus at the U. S. Animal Husbandry Experiment Farm, Beltsville, Maryland, in the fitting of growth curves.

Hendricks, W. A., A. R. Lee, and H. W. Titus. Early growth of White Leghorns, Poultry Sci. 8 (6); pp. 315-327 (1929).

Titus, H. W., and W. A. Hendricks. The Early Growth of Chickens as a Function of Feed Consumption Rather Than of Time. Conference Papers of the Fourth World's Poultry Congress, Section B (Nutrition and Rearing): pp. 285-293 (1930).

The use of such residuals leads to results which appear to give a better description of the data than when simple residuals of the type, $f(X) - Y$, are employed.

TABLE VI

Chi-square test for contingency applied to the distribution of the deviations of the type, $\frac{AX^2}{Y} - 1$. The theoretical frequencies for each compartment are given in parentheses.

Value of X	Magnitude of Deviation				
	0.000 to ± 0.019	± 0.020 to ± 0.039	± 0.040 to ± 0.059	± 0.060 and over	Total
1 to 10	4 (4.1463)	3 (3.1707)	2 (1.7073)	1 (0.9756)	10
11 to 20	4 (4.1463)	4 (3.1707)	1 (1.7073)	1 (0.9756)	10
21 to 31	4 (4.1463)	3 (3.1707)	1 (1.7073)	2 (0.9756)	10
31 to 41	5 (4.5610)	3 (3.4878)	3 (1.8780)	0 (1.0732)	11
Total	17	13	7	4	41

$X^2 = 3.8182$
$n' = 10$
$P = 0.921027$

TABLE VII

Chi-square test for contingency applied to the distribution of the deviations of the type, $1 - \frac{Y}{AX^2}$. The theoretical frequencies for each compartment are given in parentheses.

Value of X	Magnitude of Deviation				
	0.000 to ±0.019	± 0.020 to ± 0.039	± 0.040 to ±0.059	± 0.060 and over	Total
1 to 10	3 (3.9024)	4 (3.4146)	2 (1.7073)	1 (0.9756)	10
11 to 20	4 (3.9024)	3 (3.4146)	2 (1.7073)	1 (0.9756)	10
21 to 30	4 (3.9024)	3 (3.4146)	1 (1.7073)	2 (0.9756)	10
31 to 41	5 (4.2927)	4 (3.7561)	2 (1.8780)	0 (1.0732)	11
Total	16	14	7	4	41

$X^2 = 3.0984$
$n' = 10$
$P = 0.959091$

TABLE VIII

Chi-square test for contingency applied to the distribution of the deviations of the type, $\log AX^2 - \log Y$. The theoretical frequencies for each compartment are given in parentheses.

Value of X	Magnitude of Deviation				
	0.000 to ±0.009	± 0.010 to ±0.019	± 0.020 to ±0.029	± 0.030 and over	Total
1 to 10	6 (6.0976)	2 (2.4390)	2 (0.9756)	0 (0.4878)	10
11 to 20	6 (6.0976)	3 (2.4390)	0 (0.9756)	1 (0.4878)	10
21 to 30	6 (6.0976)	2 (2.4390)	1 (0.9756)	1 (0.4878)	10
31 to 41	7 (6.7073)	3 (2.6829)	1 (1.0732)	0 (0.5366)	11
Total	25	10	4	2	41

$X^2 = 4.4989$
$n' = 10$
$P = 0.872945$

EDITOR'S NOTE

It is with great pleasure that the *Annals* brings to its readers information concerning the *Nordic Statistical Journal*, edited by Dr. Thor Andersson. This publication is of great merit, and the work of its contributors compares very favorably with that found in *Biometrika* and *Metron*. Americans will do well to study carefully the contributions which Scandinavians are making to statistical methodology.

Nordic Statistical Journal

EDITED BY

THOR ANDERSSON

VOLUME 1

ARTICLES IN NORDISK STATISTISK TIDSKRIFT.

Nordisk Statistisk Tidskrift started in 1922. It is chiefly written in Nordic tongues. There are also published articles in English and German. To some articles in Nordic there are summaries in English or German. Now the chance is taken to realize the original scheme of publishing two editions, one in Nordic tongues and the other in English. The edition in non-Nordic tongues is published in English also because of the fact that the millions of descendants of the Nordic peoples, now living beyond the boundaries of the Nordic states, are mainly working in English-speaking countries.

Nordic Statistical Journal has five departments: articles, reviews of books, minor communications, bibliographical lists of Nordic statistics, and recent periodicals and new books. In general, all departments will be represented in every number.

Nordic Statistical Journal is published quarterley, the four numbers making a volume of about 640 pages. The subscription rate for a volume — post free — is 30 Swedish crowns.

Subscriptions may be sent to Nordic Statistical Journal, Stockholm, Sweden.

The subscription rate through booksellers is 35 Swedish crowns.

Editorial communications and all publications should be adressed to THOR ANDERSSON, Dr. Ph., Stockholm, Sweden.

Aftonbl. tr., Sthlm 1939.

Nordic Statistical Journal

EDITED BY

THOR ANDERSSON

VOLUME 2 PARTS 1 & 2

REPRINT AND TRANSLATION FROM NORDISK FÖRSÄKRINGSTIDSKRIFT 1930.

Nordic Statistical Journal. Volume 1. Edited by THOR ANDERSSON. Stockholm 1929. Pp. 639. Reviewed by Dr. phil. CARL BURRAU.

We, the inhabitants of the Nordic countries, are perhaps somewhat inclined to take a certain inner pride in our — as it seems to us — high civilization and to attach still more importance to ourselves in this respect during the later years, when the "Ragnarök" of the great war had devastated most of the other civilized countries and handicapped them in their competition with us. Let us hope that there are some good grounds for our selfsatisfied opinion! It is not difficult to find some facts indicating that we are right in this self-respect, even if we go to the very summits of civilization — let us think of the "Acta matematica", for instance. But if we are right, it may be very necessary for us to be on our guard against the danger of stagnation, of the standstill, where we begin to lull ourselves into the pleasant dream that our position is unshakeable, and that we may now repose on our laurels. Therefore, we must honour the persons who do not allow us to go to rest, the persons who spur us on to do our very best.

Thor Andersson is one of those whom we must honour for such an influence. In the field of statistics he seeks to be our scientific conscience. He swings his whip over our heads mercilessly and drives to activity everybody who is able to produce something, however small or great, within the field of statistics. But he is not content with that! He is not content with the achievement of having filled a long and imposing row of volumes of the "Nordisk Statistisk Tidskrift" with valuable essays and treatises written by Scandinavian as well as by leading foreign authors — all the non-scandinavian countries are now to see and feel the warmth of the light from the North. His journal is now to become an inter-

national publication, but still with an indication of its Nordic origin in its title. The first volume of the "Nordic Statistical Journal" — simultaneously forming the 8th volume of the original journal — has appeared. And it is not a trifling thing, this volume of 639 pages in great octavo! It is great in its composition, still more soaring in its purposes and ends for the future, and promising, when we consider what "the man at the wheel" has collected in these 639 pages by means of an unusual perseverance in unfailing love for the task and in spite of many — too many — external adversities.

The leading thought of the work is the same as, now soon a decennium ago, led Thor Andersson to found the *Nordisk Statistisk Tidskrift*. It is a child of the Greeks' idea of chaos and cosmos, or rather a consequent, modern continuation of this idea. Statistics is the most important means for bringing our existence over from chaos to cosmos. Statistics acquaints us with the real circumstances, and the knowledge, the real knowledge of the things, will then show how to bring things in their right places, so that the entirety becomes the arranged cosmos. But there is still much to do! We have not yet been able to elevate statistics to the rank of an observing natural science it should have, to be able to give us the real science of the things, alluded to above. In 1922 the thought was to be in the front-rank in the work for this purpose. And we have to be obliged to Thor Andersson for the strenuous work he has performed for his idea during the past years, and now it will be done on a still broader basis, i. e. for an international public, yet under Nordic leadership.

Let us study a little more closely how this new volume I seeks to perform its work in the service of the mentioned idea.

With, in a good meaning, a journalistic feeling for actualities the volume appears as a sort of jubilee-gift to *Bortkiewicz* on his sixtieth anniversary and it is therefore opened by a good full-page picture of this scientist who has given so valuable contributions to the original journal. To the reader, the following essays seem to arrange themselves into three groups which — just in order to give a name to the special groups — could be designated as *olden times*, *present times*, and *future times*.

In the group belonging to the olden times, the editor seeks to show how deeply rooted the statistical science is in the Nordic peoples by introducing a number of great men of Nordic origin, each in his way, a pioneer. These men are represented partly in full page pictures, partly in the text. It may not surprise us that none of them is a "professional statistician", for the profession is only now being created. But they belong, each in his way, to the founders of this branch. In the eightteenth century *Wargentin*, the astronomer, founded population statistics which is of fundamental importance to demographics. The essay on him is particularly well written by *Nordenmark*.

The memory of the now nearly forgotten *Eilert Sundt*, who, by his activity as a clergyman, was brought to make scientifical investigations of the society where he lives, and who gradually becomes a social-statistician of high rank, is revived both by a reprint of his peculiar essay of 1858: "On Piperviken and Ruseløkbakken (Investigations of the conditions and morals of the working-class in Christiania)" and by a scientific estimation of him ("Eilert Sundt's law") by *Rygg* who also gives an instructive account of how Sundt was disfavoured by his contemporaries, naturally in the first place by the politicians who had to do with the granting of money for his investigations! Unfortunately the politicians of the present times are not better; about that Thor Andersson himself could write a sad chapter!

Then follows *Thiele*, whose principal scientific passion, "the theory of observations", is simply the foundation of what is now more generally called mathematical statistics, and finally *Johannsen*, the investigator of heredity, who is commemorated by a picture as well as by a reprint and a translation into English of his contribution to the first volume of the original journal: "Biology and statistics".

Several other essays like those mentioned, also belong to the olden times. Thus *H. Palmströms's* essay on the first census in Norway in 1769 and that of *Thorsteinsson* on the census of Iceland in 1703.

The essays which the reader naturally refers to the "present times" are evidently caused by the editor in order to show the non-Scandinavian world the conditions in the

Nordic countries in two respects, both extremely important from a statistical point of view: the population registration and the industries, thus, firstly, how we gain our knowledge about the number and the composition of the population, and, secondly, how these people support themselves.

The editor could not have found any person more fit to write the "general" article about population registration than *Amnéus*, the director of the Oslo population register, whose institution is up to the standard and also has served as a model in many places, among others in Denmark. The condition of these matters in the different countries is further treated by the editor as far as Sweden is concerned, and as to Denmark by not less than two authors, *Bonde* and *Dalgaard*, and with regard to Finland by *Kovero*. These are very instructive essays which illustrate the importance of these things in the right way. One learns how even the "torso" (a not unjustified epithet for the arrangement introduced in Denmark, which was originally excellently planned, but which has been more than half-way broken to pieces by smallminded and short-sighted politicians, of course under the pretext of economy) of a population register, as that of Denmark, thanks to the fact that it is obligatory, gives an excellent support in many ways, among others for the 5-year censuses. One learns how deplorably far behind matters are in Wargentin's native country, where it was naturally necessary to perform registration in the large towns, but how the acclomplishment of the work is hazarded by the rather antiquated and burdensome obligatory collaboration with the clergy — and by still many other things. A survey like this, presented to an international audience is perhaps more than anything else suited to advance the population register movement, which *shall*, however, once triumph by its inner necessity.

Next the editor has intended to give a picture of the industrial statistics of the Nordic states. He has himself undertaken to treat the most important part: "the mother industry", agriculture.

Aage I. C. Jensen treats fishery, *Sköien* shipping, *Geijer* the ore resources of the Nordic countries, *Velander* the water powers of the Nordic countries, and, forestry, finally, is treated by *Aminoff* (Sweden), *Sandmo* (Norway), and *Cajander* (Finland).

To the "present times" we may also count *Storsteen's*, to us, the inhabitants of the Nordic capitals very interesting article on "The expenses of living in the Nordic capitals", a subject full of pitfalls for a less experienced statistician, but here treated with excellent fineness and with a clear prescience of the difficulties.

To the same group belongs *Thorberg:* Statistics and trade-union movement", a rather short but extremely interesting essay, not least on account of the author's position as president of the national organisation of the Swedish trade-unions. Here fall the weighty words about the social-political institution erected by the League of Nations for international labour organisation, that "the work of this organization is rendered extremely difficult by the fact that it has hitherto been almost impossible to arrive at any comparability between the statistics of the different countries".

Finally there is *Linder's:* "Some remarks on the income statistics of the census in 1920", which, according to its title, seems to belong to the present times, but which, according to its contents, is in the first place, a scientifical arithmetical example for the illustration of the applicability of Pareto's law, concluding in some wishes with regard to the future official investigations of income. This essay can therefore be said to form the transition to the last group.

When I have permitted myself to designate the third group of essays as "future", there may thereby, as a matter of fact, not be understood any paradoxical possibility of prophesying the statistics of the future. I have only wished to emphasize the editor's desire that his journal may also be one of the laboratories where the instruments for the treatment of the future statistics are created. This side of the matter has always had the editor's supreme interest; you may think only of the contributions to the previous volumes, which Bortkiewicz, Tschuprow and others have brought. We can call this side the theoretical or perhaps the mathematic-statistical one. It has been an urgent need in this volume I to show, that also we, in the Nordic countries think of this side of the matter, and among the authors of the six essays of which this group consists (besides the above mentioned contribution

by Linders) we also find all the four Nordic countries represented.

Although the chief importance of treatises of this kind would seem to fall within the realms of theory, still one of this essays, namely *Nybölle's* "Interpolation in statistics" is of rather eminent practical importance and use. Beside — or rather because of — his clear and sharp differentiation between the purely mathematical and the statistical meaning of the word interpolation and the very near connection of the last mentioned notion with that of adjustment, he here gives exclusively practical advice and instructions useful in circumstances which, so to say, belong to the every day life of the statistician. This treatise will be very welcome to many colleagues. In some opposition hereto stands *Ragnar Frisch:* "Correlation and scatter in statistical variables", in size as well as in importance one of the biggest treatises of the volume which will prick the conscience of many as it will make them clearly feel the obligation to penetrate more deeply into its contents — the author himself namely tells us that he has come "to various results, some of which are known, and others which are new, so far as I am aware" — but will easily feel frightened by the author's imposing mathematical apparatus, which is nothing less than n-dimensional vectors and appertaining matrices and orthogonal transformations. But one ought not to be frightened by these heavy implements. And this so much the less as the author presupposes no elementary knowledge in the field of vector calculation. On the contrary, he explains his whole apparatus thoroughly. There is no need of having heard words as vector or orthogonal before, and one will still be able to study this work, which, on this account has naturally become somewhat extensive (67 pages). It may be greeted with great satisfaction that one thus begins to attack the problems of theoretical statistics with such weapons. Terms such as scatter and correlation are so fundamental to science that we cannot take into use an apparatus precious enough to attack the problems contained in these terms. Here the best is not too good.

Further we meet *Jerneman*: "To the method of sampling" and *Lindeberg:* "Some remarks on the mean error of the per-

centage of correlation", essays well suited to waken respect for Nordic science abroad.

The contact with the kindred science, social economics, is in the volume attended to by *Gjermoe:* "The amplitude of industrial fluctuations", an essay covering 63 pages, which struggles with the difficult and not yet very well defined notions: times of ascent and descent, crises, fluctuations, and so on. The notion of "trend", so prominent in all the ultra modern investigations, belongs to the here used apparatus, and in a quotation — this essay is very abundant in quotations, which will be praiseworthy — we learn that it was already found by Hooker in 1901 and was defined by him as "the direction in which the variable is really moving when the oscillations are disregarded".

As rather specially belonging to the "future" we meet finally *Ångström's* brilliant essay on meteorology and statistics and have a presentiment that the latter is destined to play once the principal rôle before the former.

We will not end this review without another congratulation to Thor Andersson for having brought forth this volume I, followed by the wish and hope that his ideal struggle for the highest aims will meet with the wished for success before he grows too tired to fight against adversity. It is sadly known that the Swedish Riksdag has not granted him the necessary subsidy in spite of the fact that such recommendations as the following one could be appended to the petition:

"By the way in which Dr. Thor Andersson has edited Nordisk Statistisk Tidskrift, he has, according to our opinion, rendered great services towards the advancement of scientifical statistics and towards the spread of the knowledge of its extraordinary importance, not only for other sciences, but also, and not least, for the obtaining of a real knowledge of the social and economical structure of society, a knowledge that is necessary if the public measures of correcting social evils and of furthering industry will have the wished for effect. His name guarantees that the journal will, also in the future, hold the same prominent position as it now holds among publications of this kind.

Stockholm January 17th, 1929.

E. Phragmén. *P. G. Laurin.*"

When not even this was of any use, one does not know what to do. Judging by former experience, it would be of still less use to refer to the fact that this journal is an honour for its country (and for the Nordic states in general) as there is not like it in any of the foreign countries. For from such an argument, presented by a favourer of Eilert Sundt in the Norwegian Storting, the short-sighted politician *Jaabaek* drew the conclusion that it was surely so because such things were superfluous! The grant to Sundt was denied and he withdrew to a clergyship. Will the present politicians really dishonour themselves still more in connection with this enterprise? Let us hope that the means will be found for carrying on this great work.

Sthlm, Aftonbladets tr., 1930